الأساليب الإحصائية

الوصفية والاستدلالية في تحليل البيانات

د.عبد الله عامر الهمالي



الأساليب الإحصائية الوصفية والاستدلالية

في تحليل البيانات

فهرسة أثناء النشر إعداد إدارة الشئون الفنية - دار الكتب المصرية

الهمالي ، عبد اللّه عامر

الأساليب الإحصائية الوصفية والاستدلالية في تحليل البيانات/

تأليف: عبد اللّه عامر الهمالي

ط1ـ القاهرة: المجموعة العربية للتدريب والنشر، 2013.

665 ص: 17 × 24 سم.

الترقيم الدولي: 978-977-722-050-7

1- الاجتماع، علم الطرق الإحصائية

2- الإحصاء التحليلي

أـ العنوان

ديوي: 301,0182 رقم الإيداع: 2014/3637

الطبعة الأولى

2014



الناشر

المجموعة العربية للتدريب والنشر

8 أ شارع أحمد فخري - مدينة نصر - القاهرة - مصر

تليفاكس: 22759945 - 22739110 (00202)

الموقع الإلكتروني: www.arabgroup.net.eg

E-mail: info@arabgroup.net.eg

elarabgroup@yahoo.com

الأساليب الإحصائية الوصفية والاستدلالية في تحليل البيانات

تأليف

د. عبد اللّه عامر الهمالي

أستاذ الإحصاء الاجتماعي ومناهج البحث

كلية الآداب- جامعة بنغازي

الناشر

المجموعة العربية للتدريب والنشر



2014

بِسۡمِ ٱللَّهِ ٱلرَّحۡمَٰنِ ٱلرَّحِيمِ

﴿ وَمِنَ ٱلنَّاسِ وَٱلدَّوَآبِّ وَٱلۡأَنۡعَٰمِ مُخۡتَلِفٌ أَلۡوَٰنُهُۥ كَذَٰلِكَۗ إِنَّمَا يَخۡشَى ٱللَّهَ مِنۡ عِبَادِهِ ٱلۡعُلَمَٰٓؤُاْۗ إِنَّ ٱللَّهَ عَزِيزٌ غَفُورٌ ٢٨ ﴾

صدق الله العظيم

(سورة فاطر، الآية 28)

الإهداء

إلى أعز الأحفاد ..

أحمد

يوسف

وعبد اللّه

بداية الحاضر .. وكل المستقبل

المحتويات مُخْتصرةً

المحتويات مُفصلةً

الفصل الثامن: التطابق بين متغيرات تم قياسها على المستوى الترتيبي

الفصل التاسع: التطابق بين متغيرات تم قياسها على المستوى ذي المسافات

مقدمة

يهدف هذا الكتاب "الأساليب الإحصائية الوصفية والاستدلالية في تحليل البيانات" في المقام الأول إلى تَعلُمْ الأساليب الإحصائية واستخداماتها في تحليل البيانات بشكل يسير ما أمكن الأمر. ويتيح هذا الكتاب المرجعي فرصاً تساعد طلابنا على ممارسة الأساليب الإحصائية التي يتعلمونها من خلال الأمثلة والتوضيحات الواردة في ثنايا فصول هذا الكتاب. فليقرءوا هذه الفصول جيداً بدلاً من حفظ المعادلات. وقد عَمِدتُ بشكل علمي إلى عرض الأساليب الإحصائية في السياق التصوري الذي يشرح لماذا طوِّر هذا الأسلوب الإحصائي؟ ومتى يتم استخدامه؟ فإذا تمت قراءة فصول هذا الكتاب، والمادة العلمية التي تحتويها فصوله، وزيادة فهم المفاهيم الأساسية التي تتضمنها المعادلات الإحصائية، فالقارئ لاشك في أنه سيجد أن فهم هذه المعادلات، والطريقة التي تستخدم في سياقها، كل ذلك، بالضرورة، سيسهل له عملية فهم هذه الأساليب الإحصائية الوصفية والاستدلالية. حيث إنه لا غنى للباحث عن فهم واستيعاب كلا النوعين من هذه الأساليب كي يتمكن من تنظيم، ومعالجة، وتفسير بياناته بشكل وافٍ ومتكامل. فالفهم المستنير لهذه الأساليب الإحصائية الوصفية والاستدلالية يعتمد بدون شك، على

مدى قدرة الباحث على إلمامه واستيعابه لهذه الأساليب الإحصائية، ومن ثم قدرته على توظيفها بشكل علمي دقيق.

نأمل أن يساعد هذا الكتاب في فهم عملية تحليل البيانات، وأن يكون عاملاً مساعداً على تنفيذ الأسلوب الإحصائي الذي يتم اختياره بشكل ملائم وصحيح، وأن تُفَسَّر النتائجُ المُتَوَصَّلُ إليها بشكل دقيق يتلاءم والأسئلة البحثية المطروحة للبحث والاستقصاء.

من خلال هذا الكتاب حاولنا أيضاً أن نبين جملةً من المظاهر لتطوير القدرة الإحصائية لَدَى الدارسين والباحثين، لعل أهم هذه المظاهر، أولاً: العلاقة بين الإحصاء والبيانات، وتصميم البحث. فالتركيز على عملية تصميم البحث تشكل الأساس في عملية التحليل الإحصائي. بمعنى أن تصميم البحث الجيد أو الخطة الجيدة لجمع البيانات تعتبر عملية جوهرية للبحث الجيد. فالنتائج التي يتوصل إليها الباحث من خلال التحليلات الإحصائية لا تعتبر تحليلات جيدة، إلا إذا استندت على خطة جيدة توجه عملية جمع البيانات. ثانياً: التفسيرات الإحصائية. فالتفسيرات الإحصائية تُمثل تحدياً كبيراً للمشتغلين في مجال البحث الاجتماعي. فقدرة الباحث على تفسير البيانات الإحصائية يمكن تطويرها من خلال المواجهة والخبرة. ولتوفير المواجهة، فإننا قد حاولنا بشكل دقيق، من خلال الأمثلة الواردة في متن هذا الكتاب أن نبين المعنى الإحصائي فيما يتعلق بالسؤال الأساسي المطروح. ولتوفير الخبرة، فقد حاولت في نهاية كل فصل أن أضمن بعض الإجراءات المتعلقة باستخدام مجموعة Spss، وحاولتُ أن أبين الكيفية التي تتم بها عملية تفسير مخرجات هذه المجموعة الإحصائية. ثالثاً: ولكي تكون لدى الدارس الكفاءة في إجراء العمليات الحسابية الإحصائية، فقد ضمنت في نهاية كل فصل من فصول هذا الكتاب العديد من المسائل التي صممت أساساً للمراجعة حتى يكون الدارس على دراية بما تم طرحه من خلال هذه الفصول. رابعاً: ومن المظاهر الأخرى لتطوير القدرة الإحصائية لدى الباحثين والدارسين، فقد حاولت في سياق هذا الكتاب أن أطور من قدرة الباحثين والدارسين على قراءة التقارير والأدبيات العلمية والمهنية في مجال العلم الاجتماعي. فالشخص الواعي بالأساليب الإحصائية يمكنه استيعاب ونقد، وتقدير التقارير العلمية التي أعدت من قِبل الآخرين. هذه هي تلك المظاهر التي تشكل الإطار العام لهذا الكتاب. وفي المحصلة النهائية، فإن الهدف الأساسي من

تصميم هذا الكتاب المرجعي، هو مساعدة طلابنا، والمهتمين بمجال البحث العلمي بشكل عام، والبحث في مجال العلم الاجتماعي بشكل خاص، في فهم الأساليب الإحصائية الوصفية والاستدلالية واستخداماتها الملائمة في تحليل البيانات، ومنحهم الثقة اللازمة للتعامل مع هذه الأساليب الإحصائية واستخداماتها بشكل علمي دقيق.

محتويات الكتاب:

يتألف الكتاب من عشرين فصلاً تغطي بدءاً من الإحصاءات الوصفية وحتى الإحصاءات الاستدلالية المتقدمة. وينقسم الكتاب في ذلك إلى ستة أجزاء. يعرض الجزء الأول الإحصاءات الوصفية الأحادية: الإحصاء والبيانات، وتصميم البحث، المتغيرات والقياس، تحليل البيانات التكرارية، مقاييس النـزعة المركزية ومقاييس التشتت، المنحنى الطبيعي. أما الجزء الثاني فيتناول الإحصاءات الوصفية الثنائية: تحليل جداول التقاطع - استقصاء العلاقة بين متغيرين (الوصف الجدولي لبيانات مقاسة على المستويين الاسمي والترتيبي) - الوصف العددي للبيانات الاسمية، مقاييس التطابق الثنائية، والتطابق بين متغيرات تم قياسها على المستوى ذي المسافات والنسبي. ويتناول الجزء الثالث الإحصاءات الاستدلالية: تحليل البيانات المستخدمة من عينة واحدة، توزيعات المعاينة، التقدير، فترات الثقة، اختبار الفروض: اختبار Z لمتوسط عينة واحدة، اختبار t لمتوسط حسابي لعينة واحدة. في حين يركز الجزء الرابع على الإحصاءات الاستدلالية لعينتين أو أكثر المتمثل في: اختبار t لعينتين ذات وسطين حسابيين متساويين، فروق الدلالة: اختبار F لأكثر من وسطين، ويلقي الجزء الخامس الضوء على الإحصاءات الوصفية المتعددة من خلال التوسع في جداول التقاطع بإضافة متغيرات التحكم، الارتباط الجزئي والانحدار المتعدد. أما الجزء السادس والأخير، فيناقش الإحصاءات الاستدلالية المتمثلة في اختبار مربع كاي لفروق الدلالة لعينة واحدة لتوزيع تكراري، اختبار مربع كاي للاستقلال، اختبار التوزيع ذي الحدين لعينة واحدة، والأساليب الإحصائية للبيانات الترتيبية كاختبار مان - وتني، واختبار ولكوكسن، وكروسكال - وليز، وفريدمان.

وبعد، فإنني أضع هذا الكتاب بين أيدي الباحثين والدارسين وعامة المهتمين بالشأن

الاجتماعي، راجياً أن يلمسوا فيه نوعاً من الفائدة العلمية، وأسألهم ألا يضنّوا عليّ بأية ملاحظات قد يقترحونها أو تصويب قد يرونه فالكمال لله وحده.

وختاماً، فإنني أتقدم ببالغ الشكر لاسيما إلى كل من الأستاذ/ جمال محمود لمراجعته لمسوّدة الكتاب لغوياً والأستاذة / أسماء الشريف على عنايتها بمراجعة المسودات الأولى لهذا الكتاب، والشكر موصول إلى الآنسة / هنية العلواني على تحملها طباعة وتصحيح التجارب الطباعية في مراحلها المختلفة.

والله أدعوا أن يوفقنا جميعاً لما فيه الخير، وأن ينفع بعلمنا هذا أمتنا العربية، إنه سميع مُجيب.

د. عبد الله عامر الهمالي

جامعة بنغازي - 2011 م

الجزء الأول

الإحصاءات الوصفية الأحادية

- الفصل الأول: الإحصاء والبيانات وتصميم البحث
- الفصل الثاني: المتغيرات والقياس
- الفصل الثالث: تحليل البيانات التكرارية
- الفصل الرابع: مقاييس النزعة المركزية ومقاييس التشتت
- الفصل الخامس: المنحنى الطبيعي

الفصل الأول

الإحصاء والبيانات وتصميم البحث

مقدمة

يلعب التحليل الإحصائي الإحصائي دوراً كبيراً في العلم الاجتماعي والسلوكي، فقد لا تخلو أية مجلة علمية أو كتاب في هذين الحقلين إلاّ والاستخدامات الإحصائية حاضرة سواء كانت هذه الإحصاءات متعلقة باختبارات الدلالة أو الارتباطات أو الإحصاءات الوصفية. إن هذا الاستخدام ليس مستبعداً، فالإحصاء كأداة يستخدمها علماء العلم الاجتماعي والسلوكي هي نفس تلك الاستخدامات لدى علماء العلم الطبيعي. وباستخدام العلماء للأدوات الإحصائية فإنهم بذلك قد وصلوا إلى تحقيق أهداف عالية من حيث قدرتهم على بناء نظريات علمية قادرة على تفسير الفعل الإنساني، أو إيجاد أرضية لبناء حالة موضوعية تساعد على إيجاد برامج لخدمات اجتماعية مستمرة.

لاشك أنه بدون الإحصاء، قد نجد أنفسنا سُبَّحاً في بحر من الوقائع والأرقام، فالعقل الإنساني عادة قد لا يمكنه استيعاب هذه الحقائق والأرقام بالكيفية التي هي عليها. يجدر بنا القول بأن الباحثين في مجال العلوم الاجتماعية والسلوكية عليهم – وبشكل روتيني –

التعامل مع كم هائل من المعلومات (البيانات)، ومن هنا جاء دور الإحصاء لتلخيص وتنظيم وتحليل البيانات الرقمية بشكل مرتب وواضح يسهل فهمه بكل سهولة ويسر.

إن الاهتمام بالمعرفة الإحصائية، أصبح اليوم ضرورة ملحة، فالباحث الاجتماعي ينبغي عليه الإلمام بشكل كبير بالبرامج الإحصائية التي أصبحت اليوم مُيَسَّرة، فالباحث الُمزَوَّد بالرُّزَم الإحصائية دون فهم عميق بالكيفية التي يعمل بها الإحصاء، قد تكون خطيرة حقاً. فالبرامج الإحصائية لا تعطي الباحث أكثر مما يدخله من بيانات، ومن هنا يأتي دور الباحث المعرفي لتقييم ما تعنيه هذه المخرجات، أو بشكل أكثر دقة، ما هو المغزى الكامن وراء هذه المخرجات الإحصائية؟ فإجراء العمليات الإحصائية ليس إلا مجرد عمليات إحصائية ولكن الأمر يكمن في ما مغزى هذه العمليات الإحصائية؟ أو ما هي الدلالة وراء هذه العمليات الإحصائية؟

لاشك في أن الحاسبات قد تطورت وأن الباحث بإمكانه الحصول على التقنيات الإحصائية الأكثر تعقيداً في لمح البصر بدلاً من الأيام أو الأسابيع. وبالرغم من أن الحاسبات قد قللت من الحاجة إلى فهم آلية التحليل الإحصائي، فإنها تظل غير قادرة على اختيار أنسب أنواع الإجراء الإحصائي لمجموعة من الحالات المحددة، ولن تكون هذه الحاسبات قادرة على الإجابة على الأسئلة التي يطرحها البحث. فالإجابات تكون لدى الباحث أكثر ما تكون قيماً تم طبعه (المخرجات).

باختصار، إنَّ الاستخدام الفعال للإحصاء يحتاج إلى فهم تصوري للتحليل الإحصائي[(1)] إن الهدف من هذا الكتاب هو التركيز على الاختبارات الإحصائية واستخداماتها في مجال البحوث الاجتماعية والسلوكية. فقد خصصت فصول هذا الكتاب إلى فرد فصول مستقلة لكل اختبار من الاختبارات الإحصائية حتى لا يقع الباحث في أي إرباك أو تشويش في تطبيقه لهذا الاختبار أو ذاك. كما صُمِّم هذا الكتاب أيضاً كدليل شامل لاستخدام هذه الاختبارات الإحصائية في العلوم الاجتماعية والسلوكية.

لقد قُسِّم هذا الكتاب إلى مجموعتين من المهارات الإحصائية. أولهما: مهارات حساب وتفسير الإحصاءات الوصفية، وثانيهما: المهارات المتعلقة بحساب وتفسير الإحصاءات

الاستدلالية، فالإحصاءات الوصفية هي تلك الأعداد التي تلخص مجموعة من البيانات مع وصف أنماط في تلك البيانات. وببساطة فإن بعض الإحصاءات تصف الصفات الغالبة كالنوع، والعمر... الخ، في حين إحصاءات أخرى تصف العلاقات خاصيات النوع، الاتجاه، النوع والدخل، إن الإحصاءات الوصفية لا تقودنا إلى معرفة ما إذا كانت العينات حقيقة تمثل المجتمع الذي أخذت منه هذه العينات.

على الجانب الآخر، إن الإحصاءات الاستدلالية تساعد في تقييم ما إذا كانت التعميمات أو الاستدلالات من العينات إلى المجتمع ملائمة أم لا؟ ولفهم الإحصاءات الوصفية والإحصاءات الاستدلالية دعنا نبدأ باستعراض عملية إجراء البحث وذلك من خلال مراجعة بعض المفاهيم الأساسية لتصميم البحث الاجتماعي.

إن عملية تصميم البحث تشكل أساس التحليل الإحصائي حيث لا توجد أي كمية من الاستخدامات يمكن أن تعوض ما إذا كان هناك قصور منطقي في تلك البيانات التي تم جمعها للتحليل. بمعنى آخر، أن تصميم البحث الجيد أو الخطة لجمع البيانات تعتبر عملية جوهرية للبحث الجيد.

إن النتائج التي نتوصل إليها من التحليلات الإحصائية لا تعتبر تحليلات جيدة إلا إذا استندت على خطة توجه عملية جمع البيانات. ومن هنا رأينا أنه من الضرورة بمكان اعتماد فصل يتعلق بتصميم البحث كجزء من فصول هذا الكتاب.

إن مَعْرِفَتَنَا بالإحصاء هي أيضاً مَعْرِفَتُنَا بتصميم البحث. فالباحث الذي يعتمد بشكل كبير على اختبارات الدلالة (مثل اختبار t، واختبارات التباين... الخ) سوف يميل إلى النظر إلى المشكلات البحثية في إطار المتغيرات المستقلة والمتغيرات التابعة، المجموعات التجريبية والمجموعات الضابطة؛ ومن هنا يمكننا القول بأنه كلما كان ذا معرفة بالتقنيات الإحصائية كان أكثر مرونة في تصميم البحث. ينبغي على الباحث أن يكون قادراً على توظيف الإحصاء بشكل فعّال لتنظيم بياناته وتقيمّها وتحليلها.

إنه بدون فهم جيد لأسس التحليل الإحصائي، فالباحث سوف يكون غير قادر لبيان ما إذا كانت البيانات منطقية. وعليه يمكننا القول بأنه إذا لم تطبق التقنيات

الإحصائية بشكل ملائم، فإن البيانات المجمعة ستبقى غير ذات جدوى، ومن هنا تأتي أهمية الإحصاء الذي لا غنى عنه في العلوم الاجتماعية، فالإحصاء يزودنا بتقنيات جد مهمة في تقييم الفروض واختبار النظرية. كذلك يساعد الإحصاء علماء العلم الاجتماعي في قدرتهم على إجراء البحوث الكمية. فالبحث الكمي يقوم أساساً على تحليل المعلومات العددية أو البيانات. فالباحثون يستخدمون التقنيات الإحصائية لتنظيم ومعالجة البيانات ليتمكنوا من اختبار الفروض وتطوير النظريات أو صقلها وكذلك فهمنا للعالم الاجتماعي من أجل إصلاحه. ومن هنا يمكننا القول، بأن الباحث الذي يكون لديه خلل في تصميم بحثه، فإنه بالتالي لا يستطيع توظيف الإحصاء بشكل فعّال. فالإحصاء لا يكون بديلاً للتصورات الدقيقة والخطة المحكمة أو الاستخدام الخلاق للنظرية. فالإحصاء ليس منقذاً لتصميم بحث رديء.

عملية البحث الاجتماعي:

تمر عملية البحث الاجتماعي في العلم الاجتماعي بالمراحل التالية:

تحديد أهداف البحث:

إن أول خطوة تمر بها عملية البحث الاجتماعي أن يحدد الباحث ماذا يريد، أَيْ ما الذي يريد دراسته مبرزاً السؤال البحثي الذي سيقوده إلى الاستقصاء والبحث.

مراجعة الأدبيات المتعلقة بالمشكلة المطروحة:

ففي هذه المرحلة يُتَطَلَّبُ من الباحث أن يضع المشكلة المطروحة في الإطار العام المرتبط بالنظرية والبحث، مراجعة الدراسات السابقة حول الموضوع المطروح، هل الدراسات السابقة تتباين فيما بينها؟ هل المعلومات المتوفرة حول الموضوع يمكن للباحث أن يعالجها بشكل جيد في إطار دراسته الحالية؟

صياغة الفروض:

إن مراجعة الأدبيات والدراسات المتعلقة بالموضوع المطروح تساعد في صياغة فروض الدراسة.

القياس:

تشمل هذه الخطوة المتغيرات الأساسية للدراسة. وكيف يمكن تعريف هذه المتغيرات وقياسها؟ هل التعريفات والقياسات الواردة في هذه الدراسة تختلف عن تلك التعريفات والقياسات الواردة عند الآخرين، أم أنها تصب في نفس السياق؟ فإذا ما تم فعلياً تطوير أداة القياس (الاستبيان على سبيل المثال) أم أن الباحث سيستخدم أداة قياسية تم تطويرها من قِبَل الآخرين، فإنه من المناسب أن يضمّن هذا القياس الذي تم استخدامه من قِبل الباحث في ملاحق دراسته.

طرق جمع البيانات:

هنا ينبغي على الباحث أن يحدد ما هي الطرق المناسبة لجمع البيانات المتعلقة بالدراسة؟ هل يسعى الباحث إلى إجراء دراسة تجريبية أم مسح أم ملاحظة مباشرة أم أن الباحث يسعى للقيام بدراسة ميدانية أم أنه يريد التركيز على تحليل بيانات إحصائية تم جمعها من قِبل باحثين آخرين. - تجدر الإشارة إلى أن الباحث يمكنه استخدام أكثر من طريقة لجمع البيانات -.

تحليل البيانات:

في هذه المرحلة ينبغي على الباحث أن يشير إلى نوعية التحليل الذي سيقوم به إشارة واضحة إلى غرض ومنطق هذا التحليل. هل الباحث يرغب في تحليل بياناته بحيث لا تتجاوز الوصف؟ أم أن الباحث لديه الرغبة في تفسير لماذا تحدث الأشياء بهذه الكيفية؟ فعلى سبيل المثال، لماذا يسجل بعض الطلاب بعض الاتجاهات الإيجابية حول قضية ما دون غيرهم؟ ما هي المتغيرات التفسيرية التي يتوجب على الباحث وضعها في الاعتبار عند عملية تحليل البيانات؟ وكيف يمكن للباحث أن يعرف ما إذا كانت لديه تباينات تفسيرية دقيقة؟

الدعوة إلى التدقيق:

وذلك من خلال إتاحة الفرصة للآخرين للتعرف على البحث والتحليل. إضافة إلى

استعداد الباحث العلمي لمراجعة إجراءات عملية البحث التي تم اتباعها، البيانات التي تم جمعها والنتائج التي توصل إليها البحث.

الأسس المتعلقة بتصميم البحث:

المجتمع، عناصر المجتمع، ووحدات التحليل:

ريثما يحدد الباحث الأسئلة البحثية، عندئذ يتوجب عليه أن يقرر من هم الأشخاص المستهدفون بالدراسة للإجابة على الأسئلة التي يطرحها البحث. فالباحث عليه أن يقوم بجمع البيانات حول خصائص هؤلاء الأفراد أو المجموعات أو المؤسسات أو أي وحدة تحليل أخرى.

المجتمع:

المجتمع هو مجموعة من العناصر التي يسعى الباحث للحصول على بيانات منها أو حولها. فالعنصر هو وحدة كينونية في المجتمع. فالوحدات المفردة تؤلف مجموعة يمكن أن تكون أفراداً - مجموعات - منظمات أو مؤسسات أو أي وحدة اجتماعية منظمة (مثل القرية، المدينة) رسمية أو غير رسمية.

إن بعض المجتمعات تحتوي على مجموعة من الأفراد تمثل عنصراً لحالة منفردة من المجتمع وهي الفرد. فالمجتمع كما أشرنا، يشير إلى مجموعة من الأفراد في حين أن مجتمعات أخرى يمكن أن تحتوي على مجموعة دول، أمم، قرى، إلى آخره، والتي تتحدد بعنصر مفرد يمثل المدينة أو القرية. فالباحث عندما يرغب في دراسة معدلات الجريمة في المدن الكبرى في ليبيا، على سبيل المثال، فإن بإمكانه دراسة المجتمع (مجموعة) مثل المدن الكبرى: طرابلس، بنغازي، سبها، مصراته. أيضاً قد يرغب الباحث في دراسة معدلات الفقر في العالم، فإنه بذلك يستطيع دراسة المجتمع العالمي وهنا قد تكون وحدة التحليل عنصراً مفرداً مثل المكسيك أو أي دولة أخرى في العالم، إذاً الباحث يقوم بجمع البيانات من العناصر المكونة للمجتمع للكشف عن الأنماط في واحد أو أكثر من خصائص المجتمع ككل. إن السؤال

الذي يُطْرَحُ هنا هو: ما هي خصائص ذلك المجتمع الذي يرغب الباحث في معرفتها؟ إِنَّ الإجابة على مثل هذا السؤال تقودنا إلى الحديث عن وحدة التحليل.

وحدة التحليل:

إن وحدة التحليل هي كينونة محددة سعى الباحث لمعرفة شيء ما حولها. تجدر الإشارة إلى أنه عادة ينظر إلى عناصر المجتمع ووحدات التحليل كشيء واحد، ففي معظم الأسئلة التي تطرح في المسوح الاجتماعية الكبرى فإن الفرد يمثل وحدة التحليل في هذه المسموح. فالبيانات عادة ما تُجْمَعُ من خلال المقابلة الشخصية لهؤلاء الأفراد وجهاً لوجه لمعرفة الخصائص العامة لهؤلاء الأفراد مثل: النوع، الوضع الاجتماعي، المستوى التعليمي، الدخل... الخ، فعلى سبيل المثال، قد توجد مجموعة أخرى من الأسئلة تتعلق بالأسرة، مثل: الدخل الأسري، وعدد الأطفال في الأسرة ففي مثل هذه المتغيرات فإن وحدة التحليل هي الأسرة.

العينات وإطار المعاينة:

نادراً ما يلجأ الباحث إلى دراسة المجتمع ككل وذلك لكبر حجمه، ومن هنا يلجأ الباحث إلى اختيار عينة من ذلك المجتمع يتم اختيارها بطريقة عشوائية بحيث تمثل هذه العينة الخصائص العامة للمجتمع المدروس التي سحبت منه هذه العينة. ولكي تكون العينة ممثلة للمجتمع ينبغي على الباحث سحب العينات طبقاً للقواعد الاحتمالية. فالعينات الاحتمالية هي تلك العينات التي يتم سحبها بطريقة تتيح فرصاً متساوية أمام جميع وحدات المجتمع المدروس. وقبل سحب العينة ينبغي على الباحث أن يكون لديه إطار للمعاينة أو قائمة تحتوي على كل العناصر المكونة للمجتمع. ويتخذ إطارُ المعاينة أشكالاً مختلفةً، كالمجموعة الإحصائية والتعداد السكانيِّ، ودليل المدينة، وقوائمَ بأسماءِ الوحداتِ المطلوبِ دراستُها. وعندما يتوفر إطار المعاينة لدى الباحث عندئذ يكون بإمكانه سحب عينة عشوائية ومن التقنيات الشائعة الاستخدام لسحب عينات احتمالية، المعاينة العشوائية البسيطة، والمعاينة المنتظمة.

المعاينة العشوائية البسيطة:

وتعني المعاينة العشوائية البسيطة المعاينة التي تتيح فرصاً متساوية أمام جميع وحدات الظاهرة المدروسة. والمعاينة العشوائية بأبسط إجراءاتها يمكن من خلالها أن يكون لدى الباحث قائمةٌ حيث يتم تسجيل كل العناصر في أوراق منفصلة ثم تخلط هذه الأوراق ويختار من بينها العدد المطلوب. ويمكن للباحث أن يستخدم نفس الإجراء إلكترونياً من خلال برامج الحاسب الآلي لتوليد عينات عشوائية. أو يمكنه اللجوء إلى نفس الإجراء من خلال الجداول العشوائية بأن يختار الباحث رقماً عشوائياً من أي مكان في الجدول.

المعاينة المنتظمة ببداية عشوائية:

إن المعاينة المنتظمة ببداية عشوائية تشبه المعاينة العشوائية البسيطة، عدا أن المعاينة المنتظمة ببداية عشوائية أن فيها العنصر الأساسي يتم اختياره باستخدام جدول الأرقام العشوائية، وعندما يحدد الباحث المسافة يبدأ في الاختيار العشوائي.

عملية المعاينة متعددة المراحل:

تجدر الإشارة هنا إلى أن هناك إجراءاتٍ معاينة أكثر تعقيداً من حيث التصميم، ذلك أن هذا النوع من المعاينة يمر بعدة مراحل. فالباحث بداية عليه أن يقسم الإطارات التي من خلالها يود سحب عيناته. فالتقسيم إلى فئات يتطلب تقسيم العناصر وفقاً لخصائص معينة أو معيار معين مثل: النوع، أو الخلفية الحضرية أو الأثنية. ويتطلب من الباحث أن يقسم إطارات المعاينة ليضمن أن العينات التي يختارها ستكون ممثلة لواحد أو أكثر من الخصائص المهمة في المجتمع الذي يود دراسته. على سبيل المثال، يمكن للباحث أن يقسم الإطار إلى فئات طبقاً للنوع مولداً قوائم منفصلة للذكور والإناث، وبعد ذلك، يمكنه سحب عينة مستخدماً إما العينة العشوائية أو العينة المنتظمة من كل قائمة. إن تقسيم المجتمع إلى ذكور وإناث يسمح للباحث بأن العينة المختارة ستتضمن الذكور والإناث، نسبة لتمثيل الذكور والإناث في المجتمع ككل.

ومن استراتيجيات المعاينة الأخرى، عينة التجمعات. ويستخدم هذا النوع من المعاينة

عندما لا يكون في مقدور الباحث الحصول على إطارٍ جيدٍ وقائمةٍ للعناصر المكونة للمجتمع، وكنتيجة لذلك، ينبغي عليه في هذه الحالة أن يولد قائمة بالتجمعات أو كينونات اجتماعية منظمة تُشْبِع فضول الباحث. على سبيل المثال، إن الباحث الذي يرغب في إجراء مسح عما يريده الشباب في المجتمع الليبي، وأنه ليس بإمكانه الحصول على قائمة لكل الشباب الليبي بليبيا، فإنه في هذه الحالة ينبغي عليه أن يجد إطاراً لكل الوحدات الجغرافية التي يمكن من خلالها الوصول إلى الشباب الليبي، سواء أكانت هذه الوحدات الجغرافية مدن أو قرى أو نجوع، ومن خلال هذا الإطار يمكنه عندئذ سحب عينة عشوائية ومن هذه التجمعات مستخدماً المعاينة العشوائية أو المعاينة المنتظمة. وبعد ذلك يمكن للباحث أن يختار عشوائياً الأفراد داخل كل مجتمع للإجابة على الأسئلة التي يطرحها الباحث.

المتغيرات:

تلعب المتغيرات دوراً أساسياً في عملية البحث الاجتماعي ومن هنا ينبغي على الباحث فهم هذه المفاهيم التي تمثل الخصائص التي سَتُجْمَعُ البيانات منها. فالمتغيرات عبارة عن أي مظهر لوحدة التحليل التي يمكن أن تتباين من وحدة تحليل إلى أخرى. فإذا كانت وحدة التحليل الفرد، فإن المتغيرات يمكن أن تحتوي على خصائص مثل: العمر، النوع، عدد سنوات الدراسة، الديانة،... الخ من الخصائص. فالباحث عندما يجري بحثه فإنه في العادة يرغب في التعامل مع عدد كبير من المتغيرات فهو يقوم بجمع عدد لا بأس به من البيانات الديموغرافية والخلفية الاجتماعية للمبحوث الخاضع للدراسة.

إضافةً إلى هذه المتغيرات التي أشرنا إليها، فإن الباحث يقوم بجمع بيانات حول المتغيرات المتعلقة بالوضع الاجتماعي، والوضع المهني،... الخ، كما يمكنه أيضاً جمع بيانات حول اتجاهات الأفراد نحو قضايا معاصرة مثل الاتجاه نحو تولي المرأة مناصب إدارية عليا، أو الاتجاه نحو توزيع الثروة على الليبيين.

الفئات، والقيم، والبيانات:

يقوم الباحث بجمع بياناته عن كل متغير داخل الفئات فالسؤال حول النوع أو السؤال

حول العمر مثلاً، فإذا ما أجاب المبحوث بأنه ذكر فهنا نطلق على هذه الإجابة فئة متغير النوع أو فئة متغير العمر إذا ما حدد المبحوث عمره بالسنوات. إذاً المتغير يحتوي على الأقل فئتين (إذا تضمن أي متغير فئة واحدة فإن ذلك لن يؤدي إلى أي تباين بين المبحوثين) وعندما لا تتباين الخصائص من مبحوث لآخر، فإننا نطلق على ذلك "متغير ثابت" Constant Variable. فالباحث في بعض الأحيان يخصص أعداداً لفئات المتغيرات. يمكن للباحث أن يخصص رقم (1) لكل الإجابات التي أجابت (ذكور) عند طرح سؤال النوع. وكل واحد أجاب على سؤال النوع (أنثى) يخصص له رقم (2). وهذه الأرقام التي تم تخصيصها من قِبل الباحث لفئات المتغير يطلق عليها القيم Value.

إن البيانات المتعلقة بكل متغير تحتوي على إجابات محددة تمّ الحصول عليها من خلال السؤال البحثي المطروح. فعلى سبيل المثال، إن البيانات المتعلقة بالنوع هي إجابات محددة تم الحصول عليها عند طرح السؤال المتعلق بالنوع. والبيانات حول العمر هي تلك الإجابات المحددة التي تم الحصول عليها عند طرح السؤال حول العمر [(2)].

الفروض:

في بعض الأحيان يقوم الباحث بجمع البيانات حول متغيرات لغرض بسيط وهو وصف إجابات المبحوثين (خصائص العينة) كنسبة الذكور إلى الإناث أو نسبة المبحوثين المتزوجين أو نسبة المتعلمين في مقابل غير المتعلمين. لاشك أن الباحث قد لا يقف عند هذا الحد بوصفه لخصائص العينة، وإنما قد يرغب في معرفة العلاقة بين المتغيرات هل الذكور في المتوسط أكثر تعليماً من الإناث؟ هل صغار السن هم أكثر استعداداً لتقبل دور المرأة الجديد من الكبار؟... الخ.

تجدر الإشارة إلى أن الباحث عادة ما يكون في ذهنه بعض التخمينات حول العلاقة بين المتغيرات، وأن هذه التخمينات قد استمدها الباحث من خلال المعرفة النظرية، أو من خلال الدراسات التي أجريت من قبل أو يمكنه استخلاص هذه التخمينات من خلال التجارب الشخصية أو من خلال المشاهدات المباشرة كوّنها الباحث حول ظاهرة ما. وعندما يعبر عن هذه التخمينات أو العلاقات في عرض يصف العلاقة على الأقل بين

متغيرين، يعني هذا أن الباحث قد طوّر فرضية حول هذه العلاقة. ولنأخذ الفرض التالي كمثال على ذلك:

- الإناث أكثر ميلاً من الذكور نحو تنظيم الأسرة.

إن هذا الفرض يبين العلاقة بين النوع والاتجاه نحو تنظيم الأسرة، وبالتالي نجد أن هذه الفرضية فرضية محددة حول طبيعة هذه العلاقة. أي كيف ترتبط المتغيرات بعضها مع البعض الآخر (الإناث أكثر ميلاً من الذكور نحو تنظيم الأسرة). ولكي نختبر صحة هذه الفرضية، ينبغي علينا جمع البيانات من مجموعة الذكور ومجموعة الإناث من خلال طرح السؤال المتعلق بالاتجاه نحو تنظيم الأسرة.

أضف إلى ذلك يمكن للباحث أن يذهب بالفرض أبعد من ذلك من خلال بيان العلاقة التي تشمل متغيراً ثالثاً. فعلى سبيل المثال، يمكننا افتراض أن الرجال أكثر حرية في التعامل مع المحيط الاجتماعي الخارجي من النساء، لاسيما وأن الرجال لديهم فرصة الخروج خارج المنزل أكثر من النساء. وإذا ما صدقت هذه الفرضية، فإننا نتوقع أن نجد نسبة الرجال أكثر من نسبة النساء حيث لديهم الحرية في التعامل مع المحيط الخارجي.

المتغيرات المستقلة والمتغيرات التابعة:

من خلال العلاقة الفرضية المحددة نستطيع الحديث عن الكيفية التي ترتبط بها المتغيرات – أي أن المتغير الذي يُفْتَرَض أن له تأثيراً على المتغير الآخر – فالمتغير الذي يحدث التأثير يطلق عليه "المتغير المستقل". في حين أن المتغير الذي يتأثر يُطْلَق عليه "المتغير التابع". بكلام آخر، إن المتغير التابع هو المتغير الناتج عن تأثير المتغير المستقل. إذاً المتغير المستقل هو المتغير التفسيري الذي يُفْتَرَض أنه يؤدي إلى التباين في قيم المتغير التابع؛ وبالتالي يكون المتغير التابع النتيجة المتوقعة للمتغير التفسيري. على سبيل المثال، إذا رغب الباحث في دراسة العلاقة بين التعليم والاتجاه نحو مشاركة المرأة في الحياة العامة، فإنه قد يصل إلى نتيجة مفادها: أن الأفراد ذوي المستويات التعليمية العالية يسجلون درجات إيجابية نحو مشاركة المرأة في الحياة العامة. ويمكن توضيح هذه العلاقة بالشكل التالي:

التعليم ←——— اتجاهات إيجابية نحو مشاركة المرأة في الحياة العامة

(متغير مستقل) (متغير تابع)

مما سبق، لاحظ أننا لم نستخدم كلمة "سبب" لعرض هذه العلاقة. ولما كان متغير التعليم اُفترِضَ أن له تأثيراً على اتجاهات الأفراد نحو مشاركة المرأة في الحياة العامة، فإننا بذلك لا نستطيع أن نفترض أن تعليم شخص ما يسبب أن يسجل هذا الفرد درجات إيجابية نحو مشاركة المرأة في الحياة العامة، بالرغم من أننا في بعض الأحيان نعتقد أن العلاقة بين المتغير المستقل والمتغير التابع هي علاقة سببية أي علاقة سبب ونتيجة. وعندما يتبين لنا أن أحد الخصائص المرتبطة بأحد المتغيرات (تعليم عالٍ) عندئذٍ نميل إلى إيـجاد خاصية محددة للمتغير الثاني كذلك (مواقف إيجابية حول مشاركة المرأة في الحياة العامة). أضف إلى ذلك، أنه في بعض الأحيان يصعب علينا تحديد أي من المتغيرين يمكن اعتباره مستقلاً وأي منهما نعتبره تابعاً.

يتضح في بعض المواقف أنه لا هذا يمكن نعته بأنه متغير مستقل ولا ذاك متغير تابع عند الحديث عن العلاقة بينهما. والسؤال الذي يمكن طرحه في مثل هذه المواقف هو: ماذا نعمل؟

للإجابة على هذا السؤال يمكننا أن نسوق مجموعة أدلة عامة يمكن استخدامها في كيفية التعامل مع المتغيرات [3]:

منطق العلاقة بين المتغيرات:

ينبغي على الباحث أن يعي أن هناك سبباً منطقياً في تفكيره بأن المتغير المستقل يمكنه التأثير في المتغير التابع. هل يستطيع الباحث أن يفكر في كم التأثير الذي يحدثه أحد المتغيرات على المتغير الآخر؟ على سبيل المثال، إذا كان الباحث يرغب في معرفة العلاقة بين النوع والاتجاه نحو تنظيم الأسرة، هل بإمكانه أن يفكر في الأسباب لماذا يؤثر جنس الشخص في اتجاهه نحو تنظيم الأسرة وليس بالعكس. هل الاتجاه يؤدي إلى نتيجة الاختلاف في النوع. إن الإجابة، بطبيعة الحال، لا. إذاً، من هنا يمكننا التعامل مع النوع كمتغير مستقل، والاتجاه نحو تنظيم الأسرة كمتغير تابع [4].

الوقت:

المتغير المستقل ينبغي أن يظهر في وقت سابق على المتغير التابع. ففي المثال السابق: جنس الشخص سابق على الاتجاه نحو تنظيم الأسرة باعتبار أن الهوية النوعية لأي شخص تتطور عبر الزمن.

الخصائص النوعية في مقابل الإنجاز أو الأداء:

تعتبر الخصائص النوعية دائماً متغيرات مستقلة (الخصائص النوعية هي تلك الخصائص التي يورثها شخص ما، أو هي تلك الخصائص التي لا يمتلك الشخص القدرة على التحكم فيها). ومن أمثلة الخصائص النوعية: النوع، العرق، العمر. على سبيل المثال، قد يعامل الباحث بعض المتغيرات كمتغيرات نوعية Ascribed أو سمات موروثة Inherited Traits، بالرغم من أن هذه المتغيرات قد تتغير كالوضع الاقتصادي والاجتماعي للعائلة الأصلية (العائلة التي يولد فيها الفرد).

أما فيما يتعلق بخصائص الإنجاز والأداء، كثيراً ما يتعامل معها (ليس دائماً) كمتغيرات تابعة. وتعرف الخصائص المتعلقة بالإنجاز أو الأداء بأنها تلك الصفات attributes التي يطوّرها الفرد أو يكتسبها عند الكبر. وتكتسب هذه الخصائص من خلال المركبات: كالاختيار، والجهد والقدرة أو بفعل الوصول إلى غرض أو هدف معين، فمواقف الشخص، أو التصرفات التي يتعاطى معها، هي خصائص تم إنجازها كسنوات التعليم التي أكملها، والوضع الاجتماعي والاقتصادي الذي حققه الفرد كشخص بالغ الرشد.

ثمة مواقف يكون فيها المتغيران - في علاقتهما - مستقلَّيْن عن بعضهما البعض. ففي بعض الأحيان، نجد متغيرين، لا هذا المتغير متغير مستقل ولا ذاك المتغير متغير تابع في علاقتهما ببعض، بالرغم من أنهما ربما مرتبطان ببعضهما البعض. فعلى سبيل المثال، قد يرغب الباحثون في معرفة مواقف الأفراد الذين يقومون بدراستهم، ومع ذلك، فإن مواقف أحد الأفراد ليس بالضرورة أن تسبب أو تحدث تأثيراً على موقف شخص آخر.

إن مثل هذه العلاقات يطلق عليها علاقات متماثلة Symmetrical relationship

or associations. (سيتضح هذا المفهوم بشكل جلّي عند الحديث عن مقاييس التطابق الإحصائي في فصول هذا الكتاب). في مثل هذه الحالات لا يهم الباحثَ أيٌّ من المتغيرين يمكن التعامل معه كمتغير مستقل - باعتبار أن الآخر يكون متغيراً تابعاً - في جداول التقاطع. فالباحث في كل هذه الأحوال يمكنه أن يقرر أن أحد هذين المتغيرين متغير مستقل في علاقته بالمتغير الآخر وذلك استناداً على السؤال البحثي المطروح والمحدد الذي يحاول الباحث الإجابة عليه [5].

متغيرات التحكم:

متغيرات التحكم هي تلك المتغيرات التي يمكن أن يكون لها تأثيرُ على المتغير المستقل والمتغير التابع. فإذا افترضنا أن القدرة التحصيلية للدخل مرتبطة بالوضع العائلي فهو افتراض غير صحيح، لأن عملية الدخل عملية مرتبطة بمتغير التعليم للآباء والأبناء على حد سواء [6].

أضف إلى ذلك، لو افترضت أن النساء أكثر إيجابية نحو تنظيم الأسرة من الرجال. فإذا ما أدخلنا متغير التحكم الخلفية الحضرية والريفية، وانطلقنا من أن الخلفية الحضرية / الريفية تؤثر على العلاقة بين النوع والاتجاه نحو تنظيم الأسرة. فهذا يعني أن تأثير النوع على الاتجاه قد يختلف عند الحضريين عنه لدى الريفيين.

الخلاصة:

خلال هذا الفصل بينّا خطوات عملية البحث بَدْءًا بتحديد المشكلة، مراجعة الأدبيات المتعلقة بمشكلة البحث، صياغة الفروض، القياس، تحليل البيانات، كما تم تقديم الأسس المتعلقة بتصميم البحث وتحليل البيانات الكمية: المجتمع، سحب العينة من المجتمع، عناصر ووحدات المجتمع، والعينات، والمتغيرات، وفئات المتغيرات وقيم المتغيرات، والفروض، والمتغيرات المستقلة والمتغيرات التابعة ومتغيرات التحكم.

أسئلة للمراجعة:

1- أعط مثالاً تبين فيه المتغير المستقل والمتغير التابع في فرض يمكن بناؤه؟

2- بين من خلال دراسة سسيولوجية لبعض متغيرات التحكم؟

3- ماذا تعني بـِ:

- المجتمع الإحصائي؟
- عناصر المجتمع الإحصائي؟

1- أعطِ مثالاً يبين العلاقة بين متغيرين مع التحكم في متغير ثالث؟

2- بكلماتك الخاصة، بيّن العلاقة بين الإحصاء وعملية تصميم البحث الاجتماعي؟

3- المطلوب مراجعة أحد المجالات العلمية في مجال العلوم الاجتماعية، ثم قم باختيار بحث في مجال اهتمامك مبيناً:

- ما الإحصاءات التي تم التركيز عليها في هذه الدراسة؟
- هل هذه الدراسة قد أسست على عينة من مجتمع إحصائي؟
- كم حجم العينة؟ كيف تم اختيار هذه العينة؟
- هل النتائج يمكن تعميمها على بعض المجتمعات الإحصائية؟
- ما هي المتغيرات التي تم استخدامها في هذا البحث؟
- بيّن المتغير المستقل والمتغير التابع بهذه الدراسة؟
- حدّد مستوى القياس لكل من هذين المتغيرين؟
- ما التقنيات الإحصائية التي تم استخدامها في هذه الدراسة؟
- حاول أن تتبع التحليل الإحصائي لكي تعرف مدى فهمك لهذه التحليلات؟

الهوامش والمصادر:

أولا: الهوامش:

1- George DiEKHoff, Statistics for the Social and Behavioral Sciences: Univariate, Bivariate, Multivarite, Mc Graw Hill, INC, 1992, P.4.

2- لزيادة التوضيح يمكن الرجوع إلى: عبد الله عامر الهمالي، أسلوب البحث الاجتماعي وتقنياته، منشورات جامعة قاريونس، 2003 م.

3- J. Richard Kendrick, Social Statistics: An Introduction using SPSS for Windows Pearson Education, INC. USA, 2005, PP. 15 - 16.

4- عبد الله عامر الهمالي، أسلوب البحث الاجتماعي وتقنياته، ص ص 210 – 212.

5- J. Richard Kendrick, Social Statatics, USA, 2005 P.16.

6- عبد الله عامر الهمالي، التقنيات الإحصائية ومناهج البحث، منشورات جامعة قاريونس، 2008، ص 282.

ثانياً: المصادر:

1- Colin Robson, Real World Research, Black well Publishing, USA, 2002.

2- George DiEKHoff, Statistics for the Social and Behavioral Sciences: Univariate, Bivariate, Multivarite, Mc Graw Hill, INC, 1992.

3- Joseph F. Healey, TheEssentials of Statistics: A Tool for Social Research, Wadsworth Cengage Learning, USA. 2010.

4- J. Richard Kendrick, Social Statistics: An Introduction using SPSS for Windows Pearson Education, INC. USA, 2005.

الفصل الثاني

المتغيرات والقياس

تهتم العلوم الاجتماعية والسلوكية بالاستقصاء غير المتناهي من التساؤلات في مجالات واسعة من حقول المعرفة كالاقتصاد والاجتماع والعلوم السياسية وعلم النفس إلى آخره من حقول المعرفة في مجال العلم الاجتماعي.

وعادةً ما تطرح هذه الحقول المعرفية العديد من التساؤلات مثل:

أ- ما هو دخل الأسرة في حي معين؟

ب- ما هو معدل الجريمة؟

ج- ما هي التغيرات التي حدثت في المستوى التعليمي خلال العقود الأربعة في المجتمع العربي الليبي؟

د- ما هي اتجاهات الناس حول تولي المرأة المهام الإدارية العليا؟

وبالرغم من الاختلاف في الأسئلة المطروحة في هذا الحقل أو ذاك، إلا أن القاسم المشترك بين هذه الحقول العلمية يكمن في كونها تركز على الاستقصاء المعرفي لواحد أو أكثر من متغير. فالمتغير: "عبارة عن خاصية تجريبية تتخذ قيمتين أو أكثر فإذا كانت هذه الخاصية قابلة للتغير كماً أو نوعاً فإننا ننظر إليها كمتغير"[1]. فعلى سبيل المثال، العمر يكون متغيراً حينما يمكننا بيان الفرق العمري بين شخص وآخر.

وتجدر الإشارة هنا إلى أن الفكرة المقابلة للمتغير هي الثبات "Constant" ونعني بالثبات حالة أو خاصية لا تباين فيها بين الحالات. فعدد الدراهم في ليبيا ثابت – أي أنّ كل دينار يساوي ألف درهم، وعلى أية حال فإن معظم الأبحاث تكرس جهودها لفهم تلك المتغيرات، ولماذا يتخذ هذا المتغير خاصيات معينة لبعض الحالات دون الحالات الأخرى. إن المهمة الأساسية لتسجيل الكيفية التي يظهر بها المتغير عبر مجموعة من الحالات يطلق عليها عملية القياس أو الملاحظة [2].

فالقياس هو العملية التي من خلالها يستطيع الباحث أن يحدد ويسجل الخاصيات الممكنة للمتغير كحالة مفردة. وبمعنى آخر، فالقياس هو طريقة خاصة تتبع في قياس المتغيرات والمفاهيم الاجتماعية فمتغير النوع يأخذ قيمتين أو سمتين (ذكراً أو أنثى) ويحدد لنا القياس ما هي الفئة التي يقع فيها هذا الشخص.

إن هذه المشاهدات والقياسات للمتغير في حالة البيانات الخام المتعلقة بالبحث والتي قد نختارها من وحدات التحليل يطلق عليها حالات؛ فالحالة هي خاصية تعكس عملية القياس المرتبطة بالمتغير، فعلى سبيل المثال، إذا ما أردنا أن ندرس عملية معدلات التذكر لدى طلاب المدارس الثانوية في منطقة محددة، فالحالات المدروسة هنا هي تلك المدارس الثانوية التي مُهرت بعلامات عالية تشير إلى معدل التذكر المطبق على هذه المدارس.

في هذا المثال فإن قائمة كل المدارس الثانوية في تلك المنطقة تمثل المجتمع المستهدف، فالمجتمع هو مجموع الحالات الخاضعة لدراسة في حين تعني كلمة مجتمع في الحياة اليومية مجموع السكان أو الناس الذين يعيشون في بلد معين أو منطقة معينة، إلا أنه يمكننا الإشارة إلى أن الباحث ليس في مقدوره دراسة أفراد المجتمع المستهدف بكامله وبالتالي يكون لديه حرية اختيار جزء من هذا المجتمع وهذا الجزء هو ما يطلق عليه العينة.

فالعينة تعني مجموعة الحالات التي لا تحتوي على كل الحالات السكانية وإنما جزء من السكان. فعلى سبيل المثال، أنه من الصعوبة بمكان أن نكون قادرين على استقصاء كل المدارس في المنطقة وبالتالي فقد يلجأ الباحث لاختيار عشر مدارس لاستقصاء وقياس معدلات التذكر لهذه العينة.

ولتلخيص المفاهيم الأساسية التي أوردناها فيما سبق (المتغير، الحالات، القياس، المجتمع، العينة) انظر جدول رقم (1).

دعنا ننظر إلى واحد من الأمثلة التي تم التطرق إليها في بداية هذا الفصل والذي يمثل السؤال البحثي التالي: ما هو دخل الأسرة في حي معين؟

جدول (1): ملخص

• المتغير	الدخل.
• الحالات	الأسر.
• القياس	تحديد الدخل لهذه الأسر بعينها.
• المجتمع	كل الأسر في المنطقة عند تاريخ معين.
• العينة	من مجموع العائلات في المنطقة التي اختيرت وتم قياس الدخل من خلالها.

وبعد أن تناولنا هذه المفاهيم من خلال السؤال الذي تم طرحه الآن يمكننا أن نواجه العملية البحثية الفعلية وهي كيف يمكننا قياس دخل الأسرة؟ وما هي الأداة التي يمكننا استخدامها لكي نحدد التباين بين الأسر فيما يتعلق بالدخل؟

إنه من أجل قياس متغير لعدد من الحالات فإن التحدي الكبير الذي يواجهنا يكمن في عملية التعريفات التصورية والتعريفات الإجرائية.

التعريفات التصورية والإجرائية للمتغيرات[2]:

عند الحديث عن التعريفات التصورية والإجرائية للمتغيرات يمكننا طرح الأسئلة التالية: من أين جاءت هذه المتغيرات؟ ولماذا اختار الباحث دراسة هذه المتغيرات بشكل خاص دون غيرها؟

ولنجيب على هذه الأسئلة فإننا نود الإشارة إلى أن اختيار المتغيرات الخاضعة

للاستقصاء عادة ما تتأثر بعدد كبير ومعقد من العوامل. ويمكننا في هذا السياق أن نركز على ثلاثة عوامل وهي كالتالي:

1- الإطار النظري:

عادة ما توجد عدة طرق نظرية لتفسير الواقع من حولنا وغالباً ما تُؤْخَذ هذه التفسيرات دون جدال أي كحقيقة ثابتة بأن هذا المتغير جدير بالدراسة والاهتمام.

وتجدر الإشارة هنا إلى أنه يمكننا أن نتعامل من خلال نظرية تقليدية قائمة تهتم بمتغيرات معينة وتمثل هذه المتغيرات الأساس في نظرة هذه النظرية للعالم الواقعي. فعلى سبيل المثال حيث تركز النظرية الماركسية على الطبقة الاقتصادية كمتغير أساسي ومهم للبحث، بينما اتجاه نظري آخر قد لا يعتمد مثل هذا المتغير ويعتبره شيئاً غير ذي جدوى.

إن تحليل العالم الواقعي من خلال الطبقة الاقتصادية يعني أننا لا يمكن أن نحلله إلا في هذا الإطار دون غيره، ومثل هذه الممارسة لا يمكن أن نطلق عليها وصف ممارسة سيئة أو جيدة. فبدون نظرية - من أجل أن يكون لدينا تصور واضح حول العالم الواقعي- تصبح المشاهدات المتعلقة بالبحث غير مرتبطة ببعضها البعض بطريقة علمية ومنطقية.

وتجدر الإشارة هنا أيضاً إلى أن التصورات المسبقة التي على أساسها تم اختيار المتغيرات للاستقصاء غالباً ما تكون الأساس العلمي للبحث.

2- الأجندة المحددة مسبقا للبحث:

في بعض الأحيان قد لا يستطيع الباحثون أنفسهم تحديد السؤال البحثي والمتغيرات المطلوب دراستها أو استقصاءها، فعلى سبيل المثال قد تطلب جهة رسمية التعاقد مع بعض الباحثين لإجراء دراسة معينة أو بحث تم تحديده سلفاً من قِبل الجهة الممولة، في مثل هذا الموقف فإن الشخص أو الأشخاص يقومون بإجراء البحث بالرغم من وجود مساحة قليلة لاختيار المتغيرات المزمع استقصاءها والكيفية التي يمكن بها تعريف هذه المتغيرات باعتبار أن الباحثين يقومون ببحث لأشخاص آخرين أو جهات خاصة.

3- حب الاستطلاع يقود للبحث:

في بعض الأحيان قد لا يكون لدينا تعريف واضح للإطار النظري للاعتماد عليه في الدراسة، وكذلك لا يوجد لدينا توجيه واضح من شخص أو جهات معينة للمفاهيم الرئيسية المزمع استقصاءها وعوضاً عن ذلك فنحن نريد أن نستقصي متغيراً على أساس الاستطلاع أو على أساس تصور شعوري غير قوي قد يكون في بعض الأحيان مهماً لنشكل من خلاله متغيراً معيناً. إن مثل هذه الأحوال تكون سبباً مهماً للقيام بالبحث كنظريات أمرية.

حقاً عندما تتحول إلى حقل كلي للبحث وخاصة عندما توجد نظريات مجازفة فببساطة يمكننا الاندفاع إلى بحث ناضج جداً.

إن هذه الدوافع الثلاثة واضحة بجلاء فهي مانعة التبادل، فعلى سبيل المثال حتى وإن حددت أجندة البحث من قِبل جهات محددة فإن هذه الجهات سوف تتعامل وبكل تأكيد مع هذه الأجندة البحثية من خلال بعض الإطارات النظرية، ومهما كان الدافع فإن الاستقصاء الاجتماعي أساساً يقودنا إلى متغيرات محددة للاستقصاء. تكمن المرحلة الأولية لأي متغير في تعريفه تعريفاً تصورياً فالتعريف التصوري أو (التعريف الاسمي) لمتغير عادة ما يستخدم مصطلحات حرفية لتحديد خاصيات المتغير.

والتعريف التصوري يشبه التعريف القاموسي؛ حيث يمدنا بتعريف عامل للمتغير حتى يمكننا أن نتحصل على إحساس عام حول ما يعنيه هذا المتغير، فعلى سبيل المثال إذا ما أردنا قياس الدخل فإنه بالإمكان أن نعرف الدخل تصورياً بأنه المطلب الشرعي لشخص ما في الحصول على السلع والخدمات[3].

إنه من الوضوح إذا ما أردنا توجيه الباحثين لقياس شرعية المطالبة لشخص ما للسلع والخدمات فإنهم سيكونون في حيرة من أمرهم ولذلك فإنه يتوجب علينا تقديم مجموعة من التوجهات تسمح للباحثين فعلياً تسجيل كم من المطالب على السلع والخدمات التي سوف تتباين من شخص إلى شخص آخر، بمعنى آخر، فإن أول مرحلة لتعريف المتغير المطلوب دراسته تكمن في تعريفه على المستوى التصوري كما نحتاج إلى القواعد

والإجراءات - عمليات إجرائية - التي تسمح فعلياً بملاحظة المتغير في العالم الواقعي، والسؤال المطروح هو: ما الشيء الذي نبحث عنه لكي نبين دخل شخص ما؟ إن طرح مثل هذا السؤال يقودنا إلى مشكلة التعريفات الإجرائية، فالتعريف الإجرائي لمتغير معين يحدد الإجراءات والمحكّات التي تؤخذ لقياس ما يحتويه هذا المتغير لمجموعة من الحالات، فإذا ما أردنا مشاهدة دخل شخص ما فإن الحاجة تتطلب أن نقرر ما هي الأشياء التي ننظر إليها لكي تساعدنا لقياس هذا المتغير.

إن بيان الدخل هو مجموع كل ما يتحصل عليه الفرد بشكل مباشر مثل الأجور والمعاشات والمساعدات الاجتماعية خلال العام المنصرم، مثل هذه الأحوال هي الأساس لتعريف الدخل وبهذا التعريف الذي تَحَصَّل عليه الباحث يمكنه أن يذهب إلى العالم الواقعي ويبدأ في قياس دخول الأفراد مضافاً إليه كل النقود التي تحصّل عليها الفرد من مختلف المصادر في العام السابق.

وتعد عملية التعريف الإجرائي هي الأساس وإذا لم تكن هذه العملية هي الأساس فإن مصادر التباين سوف تحدث في البحث.

إن أي تعريف تصوري معطى باستطاعتنا تعريفه تعريفاً إجرائياً بعدة طرق مختلفة ولا يمكن النظر لواحد من هذه الطرق على أنه تعريف مطلق أو تام، على سبيل المثال، إن التعريف الإجرائي للدخل من خلال مجموع ما يتحصل عليه الشخص بشكل مباشر يترك هذا التعريف بعض المصادر الأخرى للدخل مثل الحصول على الخدمات النوعية بدلاً من الدفع النقدي المباشر لشخص.

والسؤال الذي يمكن طرحه هنا هو: ما هي المحكّات التي يمكن أن نستخدمها لكي نقرر ما إذا كان التعريف الإجرائي دقيقاً أم لا؟ إن المحك الذي يمكن استخدامه لمعرفة ما إذا كان التعريف الإجرائي صحيحاً أم لا؟ يُعْرَف في التقنيات الأدبية بمشكلة صدق المفهوم.

تصورياً يمكننا النظر إلى التعريف الإجرائي بأنه يتباين تحت المتغير الذي نفكر فيه، فميزان الحرارة الزئبقي يعد أداة جيدة لقياس التغيرات اليومية في درجة الحرارة وعندما

نتعامل مع التغيرات في المتغير "درجة الحرارة"، في وسيلة قياسه "على المزلاج الزئبقي" نجده يتغير، نفترض أننا قمنا بملء ميزان الحرارة الزئبقي بالماء بدلاً من الزئبق فإن التباينات في درجة الحرارة اليومية لن تتناسب مع التغيرات في ميزان الحرارة، وخلال يومين ربما في الحقيقة يكون هناك اختلاف في الحرارة بدون أن يكون هناك تباين يمكن أن يسجل بواسطة الأداة.

وبالرجوع إلى المثال السابق المتعلق بالدخل وبالاعتماد على التعريف الإجرائي والذي تضّمن الدفع النقدي المباشر فإنه يمكننا أن نسجل شخصين لديهما نفس الدخل وإن كانا في الحقيقة مختلفين. تصور أن شخصين قد استلما نفس كمية النقود مقابل العمل ولكن أحدهما لديه أطفال تم دفع مصاريفهم الدراسية من خلال المؤسسة التي يشتغل بها وبوضوح فإنه لا يوجد تباين بين هذين الشخصين فيما يتعلق بالدخل بالنظر إلى سيطرتهم على السلع والخدمات ولكن هذا التباين لن يسجل إذا ما اعتمدنا على التعريف الإجرائي القائم فقط على مجموع الدخل النقدي المباشر.

ولزيادة التوضيح فإذا ما انتقلنا من التعريف التصوري إلى التعريف الإجرائي لمتغير، يتوجب علينا أن نضع في الاعتبار المثال التالي:

إذا ما رغب الباحث في دراسة الجريمة فإنه يمكنه أن يعرف السلوك الإجرامي تصورياً بأنه أفعال غير قانونية من العنف ضد أعضاء آخرين من المجتمع أو ضد ممتلكاتهم، فالسؤال المطروح هنا لدى الباحث هو: كيف يمكنه أن يبين أو يعين نمط التباين في هذا المتغير؟

توجد لدينا مجموعة من التعريفات الإجرائية يمكن اعتمادها:

أ- حساب عدد المرات التي تم فيها العنف الإجرامي من السجلات الرسمية.

ب- حساب كمية الوقت التي قضاها الشخص في السجن.

ج- طرح مجموعة من الأسئلة على المبحوثين لمعرفة ما إذا ارتكبوا أية جرائم.

د- تسجيل لون شعر الرأس لدى الفرد.

وبوضوح سوف يكون من الصعوبة إثبات إن التعريف الإجرائي الأخير صحيح. إنه من غير الممكن أن نقول أن مستوى السلوك الإجرامي لدى الفرد يتغير بتغير لون شعر رأسه !.

إن التعريفات الإجرائية الأخرى تبدو أكثر التصاقاً إلى المفهوم العام للسلوك الإجرامي إلا أن كل واحد من هذه التعريفات لديه مشكلته الخاصة. فعند طرح السؤال على الأشخاص إذا ما ارتكبوا أي جريمة ربما لا يعكس هذا السؤال المقياس الصحيح لأن الناس قد لا يكونون صادقين في الإجابة حول السؤال لحساسيته أما فيما يتعلق بالسؤال حول عدد الأوقات التي تم فيها القبض على الشخص يكون غير صحيح. فقد يكون هناك شخصان في الحقيقة لديهم نفس مستوى السلوك الإجرامي إلا أن أحدهما قد يكون عدد حالات القبض المسجلة عليه أكثر بسبب انتمائه إلى الجماعات المهمشة والتي غالباً ما يكون الهدف الأساسي لدى الشرطة للقبض عليه.

إن هذا التعريف في الحقيقة يتطلب قياساً متغيراً مختلفاً عن ذلك المتغير الذي نهدف إليه. إن تحيز الشرطة أكثر إلى السلوك الإجرامي وباستخدام أيٍّ من هذه التعريفات الإجرائية لقياس إجرامية الشخص قد لا تكون مرآة صادقة لمعرفة السلوك الإجرامي على حقيقته(4).

هناك مجموعة من العوامل يمكن أن تولّد مشكلات للوصول إلى التعريف الإجرائي لمتغير بدرجة عالية من صدق المفهوم.

1- تعقد المفهوم:

إن بعض المتغيرات قد لا تكون معقدة فنوع الشخص على سبيل المثال يمكن تحديده بخاصيات فيزيقية وعلى أي حال فإن معظم المتغيرات نادراً ما تكون واضحة فقد وجدنا متغير الدخل يحتوي على جملة من الأبعاد مثل: الدفع النقدي المباشر والدفع غير المباشر أي (النوعي) وأن كل بعد من أبعاد الدخل هي متغيرات تصورية في ذاتها وتطرح مجموعة من المشكلات الإجرائية المتعلقة بها وأن كل واحد من هذه التعريفات يركز على واحد من هذه الأبعاد فالتركيز على الدفع النقدي على سبيل المثال يستبعد الأبعاد الأخرى.

2- توفر البيانات:

قد نتمكن من الحصول على تعريف إجرائي ليساعدنا على تحديد المتغير الذي نرغب في دراسته بشكل كامل، فعلى سبيل المثال يمكننا أن نفكر في عدد من حالات القبض بطريقة متصدعة لمشاهدة السلوك الإجرامي، فالباحثون قد لا يسمح لهم – نتيجة لسرية المعلومات – بمراجعة سجلات الشرطة للحصول على المعلومات المطلوبة، وعليه فإنه من غير الواضح الحصول على تعريف كامل للتعريفات الإجرائية التي يمكن استخدامها وذلك لعدم قدرتنا في الحصول على بيانات مثالية.

3- تكلفة وصعوبة الحصول على البيانات:

ولنقل إننا قادرين على مراجعة سجلات الشرطة والحصول على عدد الحالات المقبوض عليها، إن التكلفة في الحصول على ذلك يمكن أن تكون غير مسموح بها من حيث الوقت والمال وبنفس الكيفية يمكننا الشعور بأن قياساً محدداً لتلوث الماء هو مثال لتقويم تآكل النهر ولكننا نحتاج للاستعانة بخبير بتقنيات قياسية عالية بإمكانها إعاقة هذا التآكل كخيار، بدلاً من اللجوء إلى قرار مزاجي بأن الماء ضبابي يمكن أن يكون مفضلاً كقياس سريع وسهل.

4- الأخلاق:

قد تسمح لنا الشرطة بالاطلاع على المعلومات الخاصة بالقبض على الأفراد كما قد يكون لدينا المال والوقت الكافي، فهل هذا يعطينا المبرر للاطلاع على الوثائق التي ليست جزءاً من مشروع البحث؟ إن مسألة الأخلاقيات – معرفة الصواب من الخطأ – هي قضية شائكة وقد لا نستطيع إثارتها بشكل جدي هنا وإنما بطرحها كمشكلة تؤثر في التعريف الإجرائي للمتغيرات التي تظهر بشكل انتظامي في الواقع الاجتماعي المتعلق بحياة الناس.

ولهذه الأسباب أو تلك هناك جدال واسع حول مسألة الصدق Validity التي تتعلق بمشكلة التعريفات الإجرائية، وفي الحقيقة فإن هذه الجدالات تدور حول البحوث

الكمية وليست في واقع الأمر حول التقنيات أو نتائج البحث ولكن السؤال الذي يمكن طرحه هو فيما إذا كانت المتغيرات قد تم تعريفها بشكل صحيح في المقام الأول.

وتجدر الإشارة إلى أنه إذا لم نستخدم المحك الإجرائي لقياس متغير يكون ذا حساسية في الطريقة التي يتغير بها المتغير بين الحالات قد يقودنا إلى توليد نتائج غير مرضية.

فالتعريف الإجرائي لمتغير سوف يكون مألوفاً في تحديد مدى الفئات أو القيم "في بعض الأحيان نطلق عليها الدرجات" لمتغير من خلال الحالات الفردية التي يحتويها.

ونعني بهذه الفئات والقيم الحصول على مدى تباين الاحتمال الذي قد يظهر أثناء عملية القياس، فالفئات المتعلقة بالذكور والإناث يمكن تحديدها على سبيل المثال، إن المدى الكلي لتباين الحالات الفردية يمكن أن يظهر في متغير النوع وبتحديد ما لهذا المدى نكون قد أوفينا بالقواعد المتعلقة بالتعريف الإجرائي، فالتعريف الإجرائي يساعد الباحث في تحديد كل حالة في الفئة المخصصة لها في المتغير.

إن ما تناولناه في هذا الإطار هو في الحقيقة يحتوي على عنصرين أساسيين منفصلين من القياس:

أولهما: ما يطلق عليه أسس مانعة التبادل والتي تشير إلى أنه لا يمكن أن تحتوي أي حالة أكثر من قيمة واحدة لنفس المتغير، فعلى سبيل المثال، قد لا يكون عمر شخص ما ثماني عشرة سنةً، وأربعاً وستين سنة.

ثانيهما: أن القياس يجب أن يتبع أسس الشمولية التي تشير إلى أن كل حالة يمكن تصنيفها إلى مجموعة فئات، فعلى سبيل المثال، فإن المقياس لقياس الوضع العائلي يجب أن يتبع كل نمط احتمالي للوضع العائلي الذي يمكن أن يظهر وإذا لم نشمل فئة " لم يتزوج على الإطلاق " فإن هذه الحالة سوف تمدنا بتصدع خلال بعض الحالات التي تقع والتي لم نبينها في القياس[5].

تصنيف المتغيرات وفقاً لمستوى القياس:

مستويات القياس:

التعريف الإجرائي للمتغير: هو تحديد أو تعيين مدى الفئات أو القيم التي يمكن تخصيصها للحالات الفردية التي تتضمن مستوىً معيناً من القياس.

وعليه توجد لدينا أربعة مستويات للقياس وهي:

المستوى الاسمي والمستوى الترتيبي والمستوى ذو المسافات والمستوى النسبي.

وتمثل هذه المقاييس الأربعة الأساس التميزي في الإحصاء والذي من خلاله نستطيع تحديد كم من المعلومات يمكن جمعها.

في الحقيقة عندما نقرر ما هي التقنيات الإحصائية المتوفرة والتي يمكننا أن نعتمدها في تحليل البيانات فإن السؤال الذي يتبادر إلى الذهن هو: ما هو المستوى الذي على ضوئه يمكن قياس المتغير؟ وكما سنرى، هناك أشياء يمكن عملها للبيانات التي تم جمعها على مستوى المقياس ذي المسافات قد لا نستطيع عملها مع بيانات ثم جمعها على المستوى الاسمي فكلما كان المستوى من القياس عالياً كانت المعلومات المتعلقة بالمتغير كثيرة.

وفيما يلي عرضاً مفصلاً لهذه المقاييس:

1- المقياس الاسمي:

يعتبر المقياس الاسمي أدنى المقاييس ويشير هذا المقياس إلى الأعداد والرموز ويعتمد على النظام التصنيفي للأشياء أو الأشخاص، فعلى سبيل المثال نفترض أننا نرغب في معرفة الديانة للأفراد فإنه يمكننا إجرائياً تعريف ديانة لشخص ما بالرجوع إلى انتماءاته الدينية والتي يمكن تحديدها من خلال مدى الفئات التالية:

الإسلام – الهندوسية – اليهودية – المسيحية..... إلى آخره.

وتجدر الإشارة إلى أنه لضمان المقياس صحيحاً فإنه لابد أن يكون شاملاً وذلك باحتوائه على كل الفئات، إلا أن احتواء الفئات يمكن أن يكون خادعاً، وهناك طريقة

أخرى بسيطة تساعدنا على كشف المقياس الاسمي في كونه صحيحاً وذلك بإعادة ترتيب الفئات لنرى ما إذا كان الترتيب يحتفظ بمنطقيته، فعلى سبيل المثال إن أحد الترتيبات التالية لقائمة الديانات يكون صحيحاً:

المسيحية	الإسلام
الإسلام	اليهودية
اليهودية	الهندوسية
الهندوسية	المسيحية
أخرى	أخرى

وبوضوح فإن هذا الترتيب الذي ظهرت به هذه الفئات لم يتأثر بإعادة التصنيف لأن كل هذين التصنيفين لازالا يحتفظان بقاعدة مانعة التبادل والشمولية.

إن المتغير الذي يمكن قياسه على المستوى الاسمي يتباين كيفياً وليس كمياً لأن الشخص ذا الديانة المسيحية يختلف عن شخص آخر في فئة الديانة الهندوسية بالنظر إلى متغير الدين لكنهما لا يمتلكان أكثر أو أقل دين.

إن متغير الدين يمكن ترميزه بالشكل التالي:

1- الإسلام.
2- اليهودية.
3- الهندوسية.
4- المسيحية.
5- أخرى.

إن هذه الترميزات للفئات لا تحتوي على أي معنى كمي بالشكل المعتاد ولكن الأرقام التي أعطيت كان الهدف منها ببساطة هو التعريف بالفئات المختلفة ولكن ذلك لا يعبر عن العلاقة بين هذه الفئات.

تجدر الإشارة إلى أنه بإمكاننا وبكل بساطة أن نستخدم الترميز التالي لتعيين القيم الرقمية لكل فئة:

1- الإسلام.

2- اليهودية.

3- الهندوسية.

4- المسيحية.

5- أخرى تذكر..

2- المقياس الترتيبي:

بالإضافة إلى خصائصه كمقياس ترتيبي فإنه يمتلك خاصية التصنيف وكذلك فئاته مانعة التبادل وشاملة. حيث ترتب فئات هذا المقياس على حسب الخاصية لكل فئة "الترتيب الأول - الترتيب الثاني - الترتيب الثالث، الأكبر - الكبير - الأصغر..... إلى آخره".

وتساعدنا المقاييس الترتيبية في ترتيب الحالات ويتضمن هذا الترتيب ترتيب الحالات بمعنى كمي مثل: من الأدنى إلى الأعلى ومن القليل إلى الكثير أو من الأضعف إلى الأقوى ويشاع استخدام المقاييس ترتيبياً عندما يتعلق الأمر بقياس الاتجاه أو الرضا في مسوح الرأي، فعلى سبيل المثال نفترض أننا نحاول قياس الدخل فإننا يمكننا أن نحدده على المقياس التالي:

دخل منخفض - دخل متوسط - دخل عالٍ.

وتحتوي فئة ذوي الدخل المنخفض أولئك الذين يصل دخلهم إلى 15.000 ألف دينار أو أقل في حين تشمل فئة ذوي الدخل المتوسط أولئك الذين يكسبون ما بين 15.001 دينار و 50.000 ألف دينار في السنة في حين يصل ما يكسبه ذوي الفئات العليا إلى أكثر من 50.000 ألف دينار في السنة.

إن ترتيب مثل هذه الفئات ليست من مهام المقياس الاسمي ذلك المقياس الذي من خلاله يمكن تعيين الحالات إلى فئات.

وبالإضافة إلى خصائص المقياس الترتيبي في كونه يرتب الحالات فإنه يساعدنا أيضاً في معرفة شخص ما ينتمي إلى فئة ذوي الدخل المتوسط وأن لديه دخلاً أكثر من الشخص الذي يقع في فئة ذوي الدخل المنخفض.

وبطريقة أخرى فإن الشخص الذي يقع في فئة الدخل المتوسط يمكن وضعه في الترتيب الأعلى من الشخص الذي ينتمي إلى ذوي الدخل المنخفض.

إن الاختلاف بين بيانات المقياس الترتيبي وبيانات المقياس الاسمي يتمثل في كون أن الحالة في المقياس الاسمي لا تختلف من حالة إلى أخرى من حيث كونها أفضل أو أقوى أو أكبر أو أكثر قوة، فبالنظر إلى المقياس التالي:

(دخل منخفض - دخل متوسط - دخل مرتفع) نـجده يحتفظ بخصائص المقياس الترتيبي وذلك باحتفاظه بالترتيب المنطقي بين الفئات، فإذا ما قمنا بإعادة ترتيب هذه الفئات حسب خصائص المقياس الاسمي فإن ترتيب الحالات طبقاً للدخل قد يُفْقَد.

دخل متوسط - دخل عالٍ - دخل منخفض

وكما رأينا في بيانات المقياس الاسمي فإن القيم العددية يمكن أن تخصص للفئات كشكل من أشكال الاختزال في حين تصبح هذه القيم العددية في حالة المقاييس الترتيبية ضرورة للحفاظ على منطقية الترتيب وعليه فإن أحد الطريقتين لمجموعة الأرقام التالية يمكن استخدامها:

(1)	(2)	(3)
الدخل المنخفض	الدخل المتوسط	الدخل العالي
(23)	(88)	(105)
الدخل المنخفض	الدخل المتوسط	الدخل العالي

وتجدر الإشارة هنا إلى أنه يمكننا القول إلى أيٍ من نظامَيْ الترميز السابق يسمح للفئات بأن تعرف وترتب بالنسبة لبعضها البعض إلا أن هذه الأرقام في حد ذاتها ليس لديها أي دلالة كمية أكثر من كون أن وظيفتها هي الترتيب.

3- مقياس ذو المسافات والنسبي:

إذا كان المقياس الترتيبي يسمح لنا بترتيب الحالات فيما يتعلق بمتغير ما بالتالي يمكننا القول بأن شخصاً ما أو حالة ما أفضل أو أكثر قوة من الحالة الأخرى فإنه لا يسمح لنا بمعرفة الاختلاف الحقيقي بين الحالات كذلك لا يدل على مدى أو مقدار ما تمتلكه كل حالة من هذه الحالات.

وإذا ما استخدمنا المثال السابق لقياس الدخل فإننا لا نستطيع القول: كم أكثر فعلياً يكسبه شخص ما في فئة ذوي الدخل العالي مقارنة بشخص ما في فئة ذوي الدخل المتوسط، إنه قد يكون من غير السليم لنا أن نستخدم خطة الترميز الثانية المشار إليها في أعلاه، ونقول إن الشخص الذي يتحصل على دخل عالٍ يزيد بـِ 17 وحدة من دخل الشخص في الفئة ذات الدخل المتوسط، على سبيل المثال 105 – 88 – 17، ففاصل المسافة بين الفئات غير معروفة.

مثال: نفترض أننا نريد أن نقيس الدخل بطريقة أخرى وذلك من خلال طرح السؤال التالي على كل شخص لمعرفة دخل كل منهم في السنة السابقة "بالدينار الليبي" وبشكل واضح فإنه بإمكاننا أن نحدد لكل شخص الفئة التي يندرج تحتها ذلك بناءً على المعلومات التي تحصلنا عليها "القيمة المالية السنوية بالدينار" كذلك يمكننا ترتيب هؤلاء الأشخاص طبقاً لهذا المقياس "ذي المسافات" وذلك من خلال بيان أن شخصاً ما يمتلك أكثر أو أقل دخل من شخص آخر كذلك يمكننا قياس كمية الفرق في الدخل بين الحالات حيث يختلف المقياس ذو المسافات عن المقياس الاسمي والترتيبي في كونه يمتلك كمية الفرق في الدخل بين الحالات كما أن المقياس يمكّننا من أن نتحصل على الأعداد التي تعكس حقيقة القيم الكمية، أي مجموع الدينارات.

وتكمن قدرة هذا المقياس في قياس المسافات بين النقاط على المقياس ويختلف المقياس ذو المسافات عن المقياس النسبي من حيث أن المقياس النسبي يمتلك قيم الصفر المطلق التي تشير إلى عدم وجود كمية المتغير.

ومجمل القول فإنه من خلال المقياس ذي المسافات والمقياس النسبي يكون في مقدور

الباحث ليس فقط معرفة أن حالةً ما تمتلك أكبر أو أقل للمتغير تحت الدراسة من الحالة الأخرى ولكنه يستطيع القول كم هي أكثر أو أقل، فعلى سبيل المثال: إذا كان شخص ما دخله 20.000 ألف دينار ليبي فهو يمتلك أكثر بـِ 10.000 دينار ليبي من الشخص الذي يمتلك دخلاً يساوي 10.000 دينار ليبي في السنة وبهذا نستطيع أن نتعرف على المسافة بينهما، بل أكثر من ذلك فإن المسافة بين النقاط على المقياس هي قيم متساوية على المستوى العام للمدى فالفرق بين 20.000 ألف دينار ليبي و 10.000 دينار ليبي هو نفس الفرق في الدخل بين 120.000 ألف دينار ليبي و 130.000 ألف دينار ليبي.

وبوضوح فإن الأعداد أو الأرقام على المقياس ذي المسافات والنسبي ذات دلالة، حيث تشير هذه الأرقام إلى كمية قابلة للقياس. وعليه، فإن هذه الأرقام تبين قيم المتغير أو القيم المتعلقة بالمتغير. لاحظ أن "صفر" دينار يمثل حالة لا تمتلك كمية متغير الدخل وتعرف هذه الحالة بنقطة الصفر الحقيقية. وتشير إلى خاصية البيانات النسبية كنقيض للبيانات ذات المسافات المتساوية. فعلى سبيل المثال، يمكن قياس الحرارة بدرجات مئوية لا تمتلك قيمة الصفر الحقيقي حيث لا توجد نقطة. إلا أن درجة الصفر المئوية لا تشير إلى حالة "حين لا توجد حرارة" فالجو بارد ولكنه ليس بهذه البرودة، وعوضاً عن ذلك، فإن درجة الصفر المئوي تشير إلى شيء آخر - النقطة التي يتجمد فيها الماء - وعليه، فإن هذا التمييز بين المقياس النسبي وذي المسافات، فيما يتعلق بالقياس هو أمر واضح في الذاكرة، وليس بالضرورة أن نتتبع ذلك التمييز. فبإمكاننا بشكل عام أن نجري نفس التحليل الإحصائي للبيانات المقاسة على مستوى ذي المسافات، والبيانات المقاسة على المستوى النسبي.

الفرق بين المتغيرات المقولية والمتغيرات العددية:

ولكي يقرر الباحث ما إذا كانت المتغيرات التي يتعامل معها متغيرات مقولية أو متغيرات عددية، ينبغي النظر إلى الكيفية التي على ضوئها يتم جمع البيانات. فعلى سبيل المثال، إن متغير العمر يمكن قياسه على أكثر من مستوى. كالسؤال الذي يمكن طرحه: أي من الفئات العمرية التالية تمثل عمرك؟ من هنا يكون متغير العمر متغير مقولي،

ومتغير عددي، لأن البيانات التي تم جمعها في التعداد السكاني على سبيل المثال، تمَّ على ضوئها حساب العمر. السؤال الذي ينبغي طرحه هنا هو كيف لنا أن نميز بين المتغيرات المقولية والمتغيرات العددية؟

1- المتغيرات المقولية: الفرق بين المتغيرات الاسمية والترتيبية:

المتغيرات المقولية هي تلك المتغيرات التي تم جمع بياناتها من خلال فئات إجابة response categories تم تحديدها مسبقاً من قِبل الباحث.

ومن أمثلتها: المتغيرات الديموغرافية: النوع (ذكر / أنثى)، محل الإقامة (حضري / ريفي)، الحالة الزواجية (أعزب، متزوج، مطلق، أرمل)، والمتغيرات المتعلقة بالمواقف وعادة ما تجمع بيانات هذه المتغيرات في شكل فئات إجابة:

إلى أي حد يعتبر رأيك في أمور تتعلق بقرارات مهمة داخل أسرتك؟

- مهم جداً
- مهم إلى حد ما
- مهم إلى حد ما
- مهم قليلاً
- غير مهم

والمتغيرات المتعلقة بالسلوك تجمع بياناتها عادة في شكل فئات:

هناك بعض من الناس لديهم بعض الوقت للمساهمة في العمل التطوعي، بينما آخرون ليس لديهم الوقت. أنت شخصياً هل لديك الوقت للمشاركة في النشاطات التطوعية؟

- أكثر من مرة في الأسبوع.
- مرة واحدة في الأسبوع.
- بعض الأوقات شهرياً.
- مرة في الشهر.
- كل شهرين أو ثلاثة أشهر.
- على الأقل مرة في السنة.
- لا أشارك.

إن الفكرة وراء فئات الإجابة، هي أن تكون فئات الإجابة مانعة التبادل، بمعنى ألا تتداخل فئات الإجابة بحيث لا تتعدى أية إجابة الفئة المخصصة لها. قد نجد في بعض الأحيان بعض الأمثلة تتداخل فيها فئات الإجابة، مثل فئات العمر التالية:

18 – 25

25 – 35

35 – 45

+ 45 إلى آخره.

في مثل هذه الفئات يجد الباحث صعوبة في تحديد الفئة العمرية المناسبة لشخص أجاب بأن عمره 25 سنة أو 35 سنة. هذا لا يعني أن المبحوثين لن تكون لديهم صعوبة في تحديد الفئة المناسبة لمواقفهم أو سلوكهم أو أي خاصية أخرى تم قياسها. وعلى أية حال، هذا يعني أن خاصية واحدة متعلقة بالمبحوث يمكن أن تلائم أكثر من فئة للمتغير.

2- المتغيرات العددية:

يمثل المقياس ذو المسافات والنسبي المتغيرات العددية. أي تلك المتغيرات التي تجمع بياناتها كأعداد دون محاولة وضعها في فئات مسبقة من قِبل الباحث. ومن أمثلتها عدد السنوات التي قضاها شخص ما في التدريس الجامعي، عدد الساعات الأسبوعية التي يقضيها شخص ما في العملية التدريسية. لاحظ أن البيانات التي تم جمعها كمتغيرات عددية يمكن جمعها كمتغيرات مقولية (مئوية) أيضاً. ومن الأمثلة الشائعة في هذا السياق متغير الدخل، فبدلاً من أن يطرح السؤال على المبحوث: كم من النقود يتحصل عليها؟ يمكن جمع البيانات حول الدخل على المستوى العددي، كذلك يمكن للباحث أن يصنف هذه الإجابات في فئات:

أقل من 300 د.ل.

300 – 4900 د.ل.

5000 – 6900 د.ل.

7000 – 8900 د.ل.

9000 فما فوق.

المتغيرات المنفصلة والمتغيرات المتصلة:

تجدر الإشارة هنا إلى أن هناك تمييزاً آخر يؤثر في عملية القياس وهو التمييز بين المتغيرات المنفصلة والمتصلة، فالمتغير المنفصل: هو ذلك المتغير الذي لا تمكنه طبيعته من أن يأخذ جميع القيم بين حدي المتغير بل يتحرك عند أعداد معينة دون سواها فعلى سبيل المثال النوع متغير منفصل يمتلك فئتين احتماليتين هما ذكر - أنثى. ومقياس المتغيرات على المستوى ذي المسافات والنسبي فإن وحدة القياس عادة لا يمكن أن تكون قابلة للتقسيم فإذا ما نظرنا إلى عدد الأطفال في الأسرة فإنه من غير المنطق إن نقول أن لدى الأسرة 1.7 طفل باعتبار أنه لا يمكن أن يكون لنا أطفال بوحدة تقل عن طفل واحد.

وبالتالي يمكننا أن نقفز من القيمة الكلية للمتغير إلى الأخرى باعتبار أن المتغير المنفصل لا يحتوي على كسور. ولزيادة التوضيح نورد المثال التالي: لعدد السجناء لكل زنزانة، وعدد هيئات الرعايا في حي معين أيضاً عدد الحوادث الصناعية في السنة السابقة.

ويمكننا على الجانب الآخر أن نورد مثالاً متغير الرضا بالخدمات التي تقدمها المكتبة فإن مستويات الرضا بين مستخدمي المكتبة قد يكون بينها فروق ضئيلة ولذلك فإن مستخدمي المكتبة يمكن أن يتصوروا التغير بطريقة تدريجية من شخص لآخر أو لنفس الشخص عبر فترات زمنية متعاقبة هذا المثال ينطبق على المتغيرات المتصلة فالمتغير المتصل هو المتغير الذي تمكنه طبيعته من أن يأخذ جميع القيم بين حدي المتغير وإذا زاد رقم إلى الرقم التالي له يمر بكل الكسور الممكنة بينهما، فالعمر على سبيل المثال هو متغير متصل بحيث يمكن أن يقسم العمر إلى الشهور والشهور إلى أسابيع، والأسابيع إلى أيام... الخ.

إن الحد الوحيد هو بالضبط كيف نكون دقيقين فالسنوات ليست دقيقة كالشهور والشهور ليست دقيقة كالأسابيع.

نظرياً يمكننا من خلال المتغير المتصل أن نتحرك تدريجياً وبشكل سلس من واحدة من قيم المتغير إلى القيم الأخرى بدون عملية القفز.

عملياً نحن دائماً بإمكاننا القيام بعملية تقريب القياس والتعامل مع المتغير المتصل

كما لو كان متغيراً منفصلاً وهذا العلم يتيح للمقياس أن يقفز من قيمة إلى أخرى.. فعلى سبيل المثال يمكننا أن نرسخ على مقياس الرضا للخدمات التي تقدمها المكتبة بطرح أسئلة على الطلاب إذا ما كانوا راضين جداً أو راضين أو غير راضين أو غير راضين جداً من المقياس فهو مقياس منفصل بالرغم من أن المتغير في حقيقته متغير متصل.

وبنفس السياق ينطبق هذا الأمر على العمر فيمكننا قياس العمر بوحدات منفصلة مثل: السنوات أو الشهور، فالمتغير بطبيعته يزيد بطريقة متصلة.

إن استخدام المقاييس المنفصلة يُوفر لنا مجموعات منفصلة من الحالات العنقودية بمعنى آخر تعمل هذه المقاييس مثل مركز الجاذبية الذي يجذب كل التباينات الطفيفة القريبة منه في المتغير الذي يقلقنا أو يزعجنا.

فعندما نقول إن شخصين عمرهما ثمانية عشر عاماً فهما في حقيقة الأمر ربما يكونان مختلفين فيما يتعلق بالعمر إلا إذا ولدا بشكل دقيق في نفس الوقت إلا أن الفرق البسيط يمكن أن نراه بين شخص عمره ثمانية عشر عاماً وشهران وخمسة أيامٍ وساعتان وأثنتا عشرة ثانية، وشخص آخر عمره ثمانية عشر عاماً وثلاثة أشهر وأربعة عشر يوماً وسبع ساعات وثانية واحدة.

إن هذه القضية قد تكون لا علاقة لها بمشكلة البحث الذي نسعى من خلاله إلى استقصاء ومعالجة هذه الأعمار فيما يتعلق بهذا المتغير بالرغم من أنهما يختلفان حقاً [6].

إن الاختلاف البيّن بين بيانات المقياس الاسمي والترتيبي وذي المسافات والنسبي يكمن في كمية المعلومات التي يقدمها كل مستوى من هذه المستويات القياسية الأربعة.

جدول (2-1): الخصائص الأساسية لمستويات القياس

مستوى القياس	أمثلة	إجراءات القياس	العمليات الرياضية المسموح باستخدامها
الاسمى "أدنى المستويات"	النوع العرق الدين الوضع الزواجي	التصنيف إلى فئات	عدد الحالات في كل فئة مقارنة عدد الحالات في كل الفئات
الترتيبي	الطبقة الاجتماعية مقاييس الاتجاه والرأي	تصنيف الفئات + ترتيب الفئات في علاقتها ببعضها البعض	كل العمليات السابقة بالإضافة إلى الحكم بأكثر أو أقل من
ذو المسافات والنسبي "أعلى المستويات"	العمر بالسنوات عدد الأطفال الدخل	تصنيف الفئات + ترتيب الفئات إضافة إلى وصف المسافات بين الدرجات فيما يتعلق بتساوي الوحدات	كل العمليات السابقة بالإضافة إلى عمليات رياضية مثل الجمع، الطرح، الضرب... الخ

المصدر: Joseph F. HEALEY, The Essentials Statistics: A Tool for Social Research, Wadsworth Cengage Learning, USA, 2010, P.21.

يلخص لنا الجدول كمية المعلومات التي يزودنا بها كل مقياس من هذه المقاييس الأربعة، إن المهمة الأساسية تسمح لنا بالتعامل مع البيانات المجمعة على كل مقياس من هذه المقاييس فبيانات المقياس الاسمي تمتلك أقل المعلومات في حين بيانات المقياس الترتيبي تمتلك أكثر معلومات من المقياس السابق نظراً لإمكانية ترتيب الحالات في حين أن المقياس ذا المسافات والنسبي ينتزعان أكثر المعلومات لأنهما يساعدان في قياس الفرق بين القيم.

خلاصة:

حاولنا في هذا الفصل أن نناقش بشكل مبدئي الأحوال التي يجب أن نراعيها قبل استخدام المعلومات الإحصائية في التحليل، وقد تناولنا هذه القضية بشكل عام حيث إن باقي فصول هذا الكتاب ستتناول عملية التحليل وماذا نعمل بالبيانات التي يتم جمعها وبعد جمع البيانات يمكننا أن نخطو خطوة أخرى وهي عملية التحليل وأن أول خطوة في التحليل عادة ما تتعلق بوصف البيانات، وعملية وصف البيانات سنناقشها في موضعها في جزء آخر من هذا الكتاب.

كما تناولنا في هذا الفصل الفرق بين المتغيرات المقولية والمتغيرات العددية، وبشكل عام، فإن المتغيرات المقولية هي المتغيرات التي يتم جمعها من خلال فئات تم تحديدها مسبقاً من قِبل الباحث. في حين أن المتغيرات العددية هي تلك البيانات التي تم جمعها دون تحديد فئاتها مسبقاً. وتقسم المتغيرات المقولية إلى متغيرات اسمية ومتغيرات ترتيبية. فالمتغيرات الاسمية هي تلك المتغيرات التي لا تخضع فئاتها إلى ترتيب. في حين تخضع فئات المتغيرات الترتيبية إلى ترتيب. كما تناولنا في هذا الفصل العملية التي على ضوئها يتم تحديد ما إذا كان هذا المتغير متغيراً متصلاً (العمر) أو متغيراً منفصلاً (الدخل).

أسئلة للمراجعة:

1- بين مستوى القياس للمتغيرات التالية:

أ- العرق:

ليبي آسيوي

أمريكي لاتيني

أخرى تذكر

ب- الأمانة:

في طريقك إلى مدرجات الجامعة لاحظت محفظة واقعة على الأرض، وتحتوي هذه المحفظة على بعض النقود والبطاقات الشخصية: هذا الموضوع يمكن تصنيفه إلى الفئات التالية:

- إرجاع المحفظة بكل محتوياتها.
- إرجاع المحفظة ولكن الاحتفاظ بالنقود.
- لا يمكن إرجاع المحفظة على الإطلاق.

2- لقد تم طرح مجموعة من الأسئلة على مجموعة من الناس حول:

- عدد السنوات التعليمية التي تم إنجازها.
- عدد الأطفال.
- عدد الأطباء لكل 1000 / من السكان.

3- هذه مجموعة من المتغيرات تم تصنيفها في فئات:

بين ما إذا كانت متغيرات مقولية أم عددية، مقرراً بعد ذلك ما إذا كانت هذه المتغيرات اسمية أم ترتيبية؟

أ- في العادة، كم مرة تزور أقاربك؟

- أكثر من مرة في الأسبوع.
- مرة في الأسبوع.
- مرة أو مرتين في الشهر.

- مرة في السنة.
- لا أزورهم على الإطلاق.

ب- هل أنت:

عازب - متزوج - مطلق - أرمل؟

ج- ما هو المستوى الدراسي الذي تم إنجازه؟

الابتدائية - الإعدادية - الثانوية - الجامعة فما فوق - لا شيء.

د- هل توافق على تولي المرأة لموقع قيادي (رئيس وزراء مثلاً)؟

- أوافق بشدة.
- أوافق.
- الحياد.
- لا أوافق بشدة.
- أوافق.

هـ- كم دخلك الشهري؟

و- كم عدد أطفالك؟

4- بين ما إذا كانت المتغيرات التالية متغيرات منفصلة أو متصلة:

- كم عدد الأفراد الذين يعيشون في هذا المنزل؟
- كم عدد السنوات التي قضيتها في هذه المهنة؟
- هل توافق أو لا توافق على توزيع الثروة على الليبيين؟
  - أوافق بشدة.
  - أوافق بعض الشيء.
  - لا أوافق.
  - لا أوافق بشدة.

- هل تصنف نفسك حضري أم ريفي؟
- ما هو دخلك في العام الماضي؟

5- أربعة مقاييس تم تقديمها في هذا الفصل: الاسمي / الترتيبي، ذو المسافات والنسبي.

- ما هي المعلومات الإضافية التي نتحصل عليها من خلال المقياس الترتيبي مقارنة بالمقياس الاسمي؟
- ما هي المعلومات الإضافية التي نتحصل عليها من المقياس ذي المسافات إذا ما قورنَ بالمقياس الترتيبي؟
- ما هي المعلومات الإضافية التي نتحصل عليها من المقياس النسبي إذا ما قورنَ بالمقياس ذي المسافات؟

6- أجريت تجربة لمعرفة تأثير عملية القلق على التذكر. وقسمت التجربة إلى مجموعتين: مجموعة يغلب عليها القلق. والمجموعة الثانية غير ذلك أي ليست قلقة. المشاركون بشكل ثابت يتذكرون الفقرات القلقة أكثر من أولئك المشاركين غير القلقين.. بين:

- المتغير المستقل في هذه الدراسة؟
- ما هو مستوى القياس لهذا المتغير المستقل؟
- بين المتغير التابع؟
- ما هو المقياس المستخدم لقياس المتغير التابع؟

الهوامش والمصادر:

أولا الهوامش:

1- عبد الله عامر الهمالي، أسلوب البحث الاجتماعي وتقنياته، منشورات جامعة قاريونس، ط 3، 2003، ص 74.

2- يمكن الرجوع إلى عبد الله عامر الهمالي، أسلوب البحث الاجتماعي وتقنياته، المرجع السابق نفسه، في الجزء المتعلق بالتعريفات التصورية والتعريفات الإجرائية، ص ص 35 – 41.

3- George Argyrous, Statistics for Social and Health Research, With Guide to SPSS, Sage Publications, London, PP. 1 - 9.

4- Ibid, P. 7.

5- Ibid, PP. 7 - 9.

6- J. Richard Kendrick, Social Statistics: An Introduction using SPSS, Second ed, USA, 2005, PP. 35 - 48.

and George Argyrous, Statistics for Social and Health Research, op. Cit., PP. 9 - 14.

ثانيا: المصادر:

1- George Argyrous, Statistics for Social and Health Research With Guide to SPSS, Sage Publications, London, 2001 .

2- J. Richard Kendrick, Social Statistics: An Introduction using SPSS, Second ed, USA, 2005.

3- عبد الله عامر الهمالي، أسلوب البحث الاجتماعي وتقنياته، منشورات جامعة قاريونس، 2003م.

4- ــــــــــــــــ، التقنيات الإحصائية ومناهج البحث، منشورات جامعة قاريونس، 2008م.

الفصل الثالث

تحليل البيانات التكرارية

مقدمة

نركز في هذا الفصل على الإجراءات المستخدمة في تلخيص المعلومات الهائلة التي يتحصل عليها الباحث من خلال إجراء بحثه. ويطلق على هذه الإجراءات، الإحصاءات الوصفية. وتشير الإحصاءات الوصفية إلى الأعداد والتمثيل البياني، وتقنيات الجدولة لتنظيم البيانات وعرضها وتحليلها.

ومن المميزات الكبرى للإحصاءات الوصفية أنها تلخص ذلك الكم الهائل من البيانات حتى يسهل قراءتها بشكل يَسِير. وباختزال كمية البيانات الكبيرة إلى إحصاءات قليلة أو في شكل تمثيل بياني أو جداول تكرارية؛ فإن نتائج البحث سَتُعْرَضُ بشكل مختصر ودقيق.

نفترض أننا قمنا بإجراء مسح اجتماعي لغرض الحصول على بيانات حول دخل عشرين أسرة. وجاءت نتائج هذا المسح كالتالي [1]:

520	462	400	360	0
560	470	420	375	0
700	475	425	400	250
1020	562	450	400	300

إن هذا الترتيب لقياس الدخل يطلق عليه توزيع، وقد تم تقديم هذه البيانات في شكل بيانات خام Raw data كنتائج لهذا المسح. لاحظ أنه إذا عُرِضَت البيانات في شكل بيانات خام، فإن ذلك لا يعطينا أي معنى دال للمتغير الذي نسعى لدراسته. فالبيانات حول دخل هذه الأسرة وبالطريقة التي عُرِضَتْ بها هذه الدخول لهذه الأسر لا معنى لها.

ثمة طريقة أخرى يمكننا اعتمادها في عرض البيانات. وهذه الطريقة يمكن من خلالها عرض هذه البيانات في شكل يمكن معه توليد أرقام قليلة - إحصاء - بحيث تكون هذه المعلومات ذات صلة بما تحتويها البيانات الخام. فالإحصاءات الوصفية عادة ما تمدنا بصورة مهمة للتوزيع الذي لم يكن واضحاً عندما عُرِضَت البيانات في شكلها الخام.

إن أحد هذه الصور الواضحة التي سنتناولها في الفصول اللاحقة من هذا الكتاب ترتبط بفكرة المعدل. فعلى سبيل المثال، يمكننا حساب رقم مفرد لمتوسط الدخل، ويُقَدم هذا الرقم المفرد كجزء من نتائج البحث.

إن مقياس المتوسط الذي تم اختياره لا يصور لنا كل البيانات التي تحتويها البيانات الأولية، وإنما يعطي هذا المتوسط فكرةً عامةً على ما تكون عليه هذه الحالات العشرون، مع السماح للباحث بإيجاد بعض التفسيرات الدّالة.

أنواع الإحصاءات الوصفية:

ما أشرنا إليه حتى الآن هو تقديمنا لفكرة المعدل كصورة مهمة لتوزيع الدرجات التي يرغب الباحث في معرفتها، وبمصطلحات تقنية يطلق علي هذا الأمر مقاييس النزعة

المركزية، ومقاييس التشتت كإحصاءات وصفية. ويطلق على هذين النوعين من الإحصاءات الوصفية: التقنيات العددية Numerical Technigues لوصف البيانات لأنها تتضمن استخدام المعادلات الرياضية لإجراء العمليات الحسابية من البيانات الخام (ستتناول هذه التقنيات العددية في الفصول اللاحقة من هذا الكتاب).

إضافة إلى هذه التقنيات، توجد جملة من الطرق الأخرى التي يمكن من خلالها عرض البيانات ليسهل قراءتها. ومن هذه الطرق المستخدمة في عرض البيانات طريقة بناء الجداول التي تصف توزيع الحالات عبر مدى من القيم لمتغير ما. وكذلك الإحصاءات الثنائية الوصفية مثل: الجداول الثنائية Bivariate Tables، ورسم شكل الانتشار Scatter Plots، ومقاييس التطابق Measures of association. ويمكن تلخيص هذه الطرق المتعددة لوصف البيانات في الجدول التالي:

جدول (3-1): أنماط الإحصاء الوصفي

النمط Type	الوظيفة Function
• الجداول	تمدنا بتوزيع تكراري لمتغير
• التمثيل البياني	يمدنا بعرض بصري لتوزيع متغير ما
• مقاييس النزعة المركزية	حساب معدل الدرجة لتوزيع
• مقاييس التشتت	حساب تباين الدرجات لتوزيع
• مقاييس التطابق	تشير إلى وجود واتجاه وقوة العلاقة بين متغيرين أو أكثر

المصدر: George Argyrous, Statistics for Social and Healh Research, op. Cit, P. 40.

آخذين في الاعتبار أن هناك عدداً كبيراً من الإحصاءات الوصفية المتوفرة. والسؤال الذي يمكن طرحه هنا هو كيف يمكن للباحث أن يقرر ما هي الإحصاءات الوصفية التي يمكن استخدامها في سياق بحث محدد؟ قد يلجأ الباحث في بعض الأحيان إلى تلخيص بياناته من خلال استخدام الجداول. وفي أحيان أخرى إلى استخدام المعدلات و/أو

مقاييس التشتت. بمعنى آخر، إن اختيار الإحصاءات الوصفية لتلخيص البيانات تعتمد بالدرجة الأولى على السؤال البحثي المطروح. وبالرغم من هذا كله، هناك بعض العوامل المحددة تستخدم كطرق إرشادية تساعد الباحث في اختيار الإحصاءات الوصفية. فعلى سبيل المثال، يعتبر مستوى القياس ذا أهمية بالغة في تحديد نوعية الإحصاءات الوصفية المستخدمة الذي سنتناوله بالتفصيل في موضعه في الفصول اللاحقة.

الجداول التكرارية:

الجداول التكرارية هي تلك الجداول التي تلخص إجابات المبحوثين في كل فئة من فئات المتغير المدروس. وببساطة هي عبارة عن قوائم لتوزيع المبحوثين (عدد الإجابات) في كل فئة من فئات المتغير. وتعتبر الجداول التكرارية طريقة شائعة الاستخدام في وصف البيانات للحصول على نتائج بحثية دالّة. أي أنها جداول تلخص توزيع متغير ما وذلك بحصر الحالات في كل فئة من فئات هذا المتغير. كما تساعد هذه الجداول أيضاً في تنظيم البيانات وتحليلها، وبالتالي تمثل الجداول التكرارية الخطوة الأولى في أي تحليل إحصائي. ومن خصائص التوزيع التكراري أنه يمكن استخدامه لكل أنواع المتغيرات، سواء كانت متغيرات عددية أو متغيرات مقولية، متصلة أم منفصلة.

ويستخدم الباحثون الجداول التكرارية لتلخيص الخصائص الديموغرافية مثل: العمر، النوع، الخلفية الحضارية، الحالة الاجتماعية... الخ. وذلك عندما يود الباحث استخدام هذه المتغيرات الديموغرافية لتحليل الاتجاهات والمواقف لأولئك المبحوثين الذين يرغب الباحث في دراستهم. كما تساعد الجداول التكرارية على معرفة الأنماط العامة للبيانات. كما تساعد التوزيعات التكرارية في تقييم تشتت الإجابات للمتغير ليقرر الباحث ما إذا كانت هذه التوزيعات توزيعات متغايرة أم توزيعات متجانسة. فالتوزيعات المتغايرة هي تلك التوزيعات التي يكون فيها توزيع المبحوثين حول المتغير يكاد يكون توزيعاً متساوياً، مشتتاً، أو منتشراً عبر كل فئة من فئات المتغير. ففي التوزيع المتغاير الكامل تكون فيه كل فئة من فئات المتغير لديها نفس (نسبة) عدد المبحوثين. فعلى سبيل المثال، متغير النوع تكون توزيعاته متغايرة 50 % من الذكور، و 50% من الإناث.

تجدر الإشارة إلى أنه كلما زاد انحراف التوزيع أكثر من تساوي تشتت المبحوثين عبر فئات المتغير، قلت عملية التغاير.

وعلى الجانب الآخر، فالتوزيعات المتجانسة هي تلك التوزيعات التي يتركز فيها المبحوثون في فئات قليلة فقط للمتغير. أي أن توزيع التجانس الكامل يعني أن كل المبحوثين يتمركزون في فئة واحدة للمتغير مثل: النوع: ذكور (100) والإناث (0)، الوضع الاجتماعي: منخفض (0)، طبقة عاملة (100)، وسطى (0)، عليا (0).

أشكال التوزيعات التكرارية(*) :

تتخذ التوزيعات التكرارية عدة أشكال منها:

1- جداول البيانات المدرجة Listed data tables
2- جداول التوزيعات التكرارية البسيطة Simple frequency tables
3- جداول التوزيعات التكرارية النسبية Relative frequency tables
4- جداول التوزيعات التكرارية المتجمعة Cumulative frequency tables

1- جداول البيانات المدرجة:

ولتوضيح جداول البيانات المدرجة نفترض أننا تحصلنا على النتائج التالية من خلال مسح افتراضي لعدد 20 حالة. وأن الباحث يرغب في التركيز على ثلاثة متغيرات هي النوع (تم قياسه على المستوى الأسمي)، والرضا المهني (ترتيبي)، وأخيراً متغير العمر (ذو المسافات المتساوية والنسبي). ويوضح الجدول (3 – 2) هذه البيانات الافتراضية.

(*) يمكن للباحث أن يبني هذه الجداول يدوياً إذا كان عدد الحالات التي يتعامل معها صغيرة أو باستخدام برنامج Spss.

جدول (3 ـ 2): نتائج المسح الافتراضي لعدد 20 حالة

رقم الحالة	النوع	الرضا المهني *	العمر
1	ذكر	متمتع بصحة جيدة جداً (3)	18
2	ذكر	متمتع بصحة جيدة جداً (3)	21
3	أنثى	تتمتع بصحة جيدة (2)	20
4	ذكر	غير متمتع بصحة جيدة (1)	18
5	أنثى	متمتعة بصحة جيدة جداً (3)	19
6	ذكر	غير متمتع بصحة جيدة (1)	18
7	أنثى	متمتعة بصحة جيدة جداً (3)	22
8	ذكر	متمتع بصحة جيدة جداً (3)	19
9	أنثى	متمتعة بصحة جيدة جداً (3)	18
10	ذكر	متمتع بصحة جيدة جداً (3)	20
11	ذكر	غير متمتع بصحة جيدة (1)	18
12	أنثى	متمتعة بصحة جيدة جداً (3)	19
13	ذكر	متمتع بصحة جيدة جداً (3)	22
14	ذكر	متمتع بصحة جيدة جداً (3)	19
15	أنثى	غير متمتعة بصحة جيدة (1)	20
16	أنثى	متمتعة بصحة جيدة (2)	18
17	ذكر	غير متمتع بصحة جيدة (1)	21
18	أنثى	غير متمتعة بصحة جيدة (1)	19
19	ذكر	متمتع بصحة جيدة (2)	18
20	ذكر	متمتع بصحة جيدة جداً (3)	20

* راضٍ جداً (3)، راضٍ (2)، غير راضٍ (1).

المصدر: George Argyrous, op.cit,P.41.

يطلق على مثل هذا الجدول، جدول البيانات المدرجة في قائمة، باعتبار أن كل درجة لكل حالة لكل متغير تم إدراجه منفصلاً. أي كل حالة في فئة واحدة فقط دون غيرها.

إن مثل هذا النوع من الجداول به عدد كبير من الصفوف بقدر الحالات، وعدد كبير من الأعمدة بقدر المتغيرات التي يتم مشاهدتها. إن مثل هذا الشكل من عرض البيانات لا يكشف لنا عن معلومات مفيدة، خاصة عندما يكون لدينا عدد كبير من الحالات يصبح التعامل مع هذا العدد من الحالات الكبيرة غير عملي. فقد يكون في مقدور الباحث أن يتعامل مع عدد محدود من الحالات، ولكن الأمر لن يصبح ميسوراً إذا زاد عدد الحالات عن عشرين. فإذا تصورنا أنه لدينا مسح لعدد 2000 شخص بدلاً من عشرين، فإنه ليس في إمكاننا بناء جدول بيانات مُدرجة (المساحة). إن الميزة المتعلقة بالبيانات الخام لكل حالة منفصلة، إنها تتيح للباحث إمكانية إجراء عدة حسابات متعلقة بالإحصاء الوصفي.

2- جداول التوزيعات التكرارية البسيطة:

تُسْتَخْدَمُ جداولُ التوزيعات التكرارية البسيطة لوصف البيانات المتعلقة بقيم متغير ما تحت الدراسة والمقاس على المستوى الاسمي والترتيبي وذي المسافات والنسبي. ومن البيانات الخام التي تم تقديمها أعلاه يمكننا بناء مجموعة من الجداول المنفصلة لكل متغير من المتغيرات الثلاثة.

جدول (3-3): جنس المبحوثين

النوع (X)	التكرار (F)
ذكور	12
إناث	8
المجموع	20

جدول (3ـ4): الرضا الصحي للمبحوثين

الرضا الصحي (X)	التكرار (F)
غير راضٍ	7
راضٍ	5
راضٍ جداً	8
المجموع	20

جدول (3ـ5): عمر المبحوثين

العمر بالسنوات (X)	التكرار (F)
18	7
19	5
20	4
21	2
22	2
المجموع	20

تشير هذه الجداول إلى مجرد البناء الأدنى الذي ينبغي أن تحتويها كل الجداول التكرارية بغض النظر عن المتغير ومستوى قياسه. ويتمثل الحد الأدنى لبناء الجداول التكرارية البسيطة في:

- أن يكون للجدول عنوان يحدد بوضوح أسماء الفئات.
- حصر الوقائع المتعلقة بكل فئة من فئات المتغير موضوع الدراسة بحيث تكون شاملة ومانعة التبادل.
- الحصر الكلي للحالات (N).

ويلاحظ في الجدول المتعلق بتوزيع المبحوثين حسب النوع أن الصف الأول يحتوي على الذكور، في حين يحتوي الصف الثاني على الإناث. وهذا التصنيف تصنيف اعتباطي يتم استخدامه في حالة المقياس الاسمي. ويمكننا أن نعيد تنظيم هذه الفئات بأي طريقة أخرى يختارها الباحث. ولقد تم وضع الذكور أولاً باعتبار أن ذلك ممارسة عامة في المقياس الاسمي على أن تنظم الصفوف بكيفية تكون الفئة ذات التوزيع الأعلى في الصف الأول. والفئة التي تليها من حيث العدد في الصف الثاني وهكذا... ويرجع السبب في هذا الترتيب إلى "فئة نموذج" Model Category الذي يستخدم في تحليل توزيعات المتغير الاسمي. في حين أن توزيع المبحوثين حسب درجة الرضا الصحي، والعمر فإن فئاتهما تخضع إلى قاعدة صارمة وفقاً للحقيقة أن التعامل معهما وفقاً للمقياس الترتيبي وذي المسافات والنسبي. حيث إن الإجراء المتبع في هذه المستويات من القياس، عامة، يبدأ بأقل قيمة في التوزيع ثم ترتقي هذه القيم تدريجياً في القياس إلى الأسفل بحيث نتحرك في الجدول صفاً بصف[2].

3- جداول التكرارات النسبية:

تعبر الجداول التكرارية النسبية عن عدد الحالات داخل كل قيمة من قيم المتغير كنسبة مئوية، أو تناسب من المجموع الكلي للحالات. ولكي نولد جدولاً تكرارياً نسبياً من البيانات السابقة التي تعرضنا لها، ينبغي على الباحث أن يلم بتقنيات النسب المئوية والتناسب.

أ- النسب المئوية:

النسب المئوية عبارة عن إحصاءات تقيس العدد الكلي للحالات استناداً لقاعدة قيمتها (100). والمعادلة التالية تُسْتَخْدَمُ لحساب النسبة المئوية:

$$\% = \frac{F}{N} \times 100$$

حيث إن: F يشير إلى التكرار وعدد الحالات في أي قيمة معينة.

n تشير إلى العدد الكلي للحالات لجميع الفئات.

ولتوضيح العملية التي يتم بها حساب النسب المئوية، نورد المثال التالي:

نفترض أن لدينا إحصاءاتٍ تشير إلى الوضع الاجتماعي لمجموعة من الأفراد يبلغ عددهم 133 فرداً، وكانت أوضاعهم الاجتماعية كالتالي:

عدد المتزوجين 63 حالة، والمطلقين 19 حالة، والمخطوبين 27 حالة، والأرامل 13 حالة، وآخرون 11 حالة. فإن النسب المئوية لهؤلاء تحسب كما يلي [3]:

- المتزوجون = $\frac{63}{133} \times 100 = 47.3\ \%$
- المطلقون = $\frac{19}{133} \times 100 = 14.3\ \%$
- المخطوبون = $\frac{27}{133} \times 100 = 20.3\ \%$
- الأرامل = $\frac{13}{133} \times 100 = 9.8\ \%$
- آخرون = $\frac{11}{133} \times 100 = 84.3\ \%$

ب- التناسب Proportions:

يوفر التناسب والنسب المئوية إطاراً مرجعياً بطريقة تجعل من البيانات أكثر وضوحاً. إلا أن الشيء الذي يمكن ملاحظته عند التعامل مع التناسب (P)، يستند على أساس قاعدة واحد صحيح، بدلاً من مائة كما هو الحال في النسبة المئوية.

$$P = \frac{F}{N}$$

إن النتيجة التي سنتحصل عليها من خلال التناسب يمكن التعبير عنها من خلال التعبير العشري decimal. ففي المثال السابق:

- المتزوجون = $\frac{63}{133} = 0.473$

- المطلقون = $0.143 = \frac{19}{133}$
- المخطوبون = $0.203 = \frac{27}{133}$
- الأرامل = $0,098 = \frac{13}{133}$
- آخرون = $0.083 = \frac{11}{133}$

وتجدر الإشارة هنا - بشكل عام - إلى أن التعامل مع النسب المئوية أسهل منه مقارنة بالتعامل مع التناسب؛ ولسبب أو آخر، فإن الناس يفضلون التعامل مع الرقم الكلي بدلاً من العشريات. إلا أننا وفي ثنايا هذا الكتاب سنتعامل بشكل كبير مع التناسب ومن هنا ينبغي على القارئ أن يتعلم العلاقة البسيطة بين التناسب والنسب المئوية المألوفة (4).

ولتحويل التناسب إلى ما يقابله من قيمة مئوية يكون بتحريك النقطة العشرية مكانين إلى اليمين (هذه العملية تكون تماماً لو ضربنا الرقم في مائة). ولكي تُحَوَّلَ النسبة المئوية إلى ما يقابلها من تناسب يتم ذلك بتحريك النقطة العشرية إلى اليسار (هذه العملية تعطي نفس النتيجة إذا تم تقسيم النسبة المئوية على مائة).

ثمة بعض المحاذير ينبغي على الباحث أن يضعها في الاعتبار عند التعامل مع التناسب والنسب المئوية. أولهما عليه النظر عندما يواجه النسب أو التناسب هو مجموع الصف الذي تم حساب النسب والتناسب منه. ذلك أن النسب المئوية والتناسب تستخدم في بعض الأحيان لإخفاء فروق شديدة في الحجم المطلق. فالزيادة على سبيل المثال، في معدلات البطالة من 10 % إلى 10.5 % لا تبدو زيادة شديدة بالمفهوم الإحصائي، في حين أن نسبة 0.5 تمثل 35.000 شخصاً، وبالتالي تعتبر هذه الزيادة بمفهوم الوضع الاقتصادي والاجتماعي زيادة كبيرة.

وبشكل معاكس، فإن التغير الكبير في أرقام النسب المئوية يمكن أن يكون أمراً عادياً

عند التعامل بأرقام مطلقة. وعلى سبيل المثال: فإن عدد الناس الذين حضروا الاجتماع الأخير يزيد بنسبة 150 في المائة أكبر من نسبة أولئك الذين حضروا الاجتماع السابق ولكن إذا كان هذا فعلياً يرجع إلى حضور خمسة أفراد في الاجتماع الأخير بدلاً من شخصين في الاجتماع السابق، فإنه قليلاً ما يكون ازدياداً شديداً. فعند التعامل مع أرقام مطلقة صغيرة، فإن إضافات صغيرة إما للمجموع أو للفئات المكونة لهذا المجموع سوف تؤثر كثيراً في الرقم النسبي المحسوب[5].

وعليه، عند التعامل مع عدد صغير من الحالات أي أقل من عشرين، ففي هذه الحالة يفضل الاعتماد على التكرارات الفعلية بدلاً من النسب المئوية والتناسب [6].

بعد أن تعرفنا على النسب المئوية والتناسب واستخدامها في بناء جداول التوزيعات التكرارية النسبية للبيانات التي تم استخدامها سابقاً. يمكننا أن نضيف للجدول لكل متغير، عموداً يوضح النسبة المئوية (أو التناسب) للحالات التي تقع في كل فئة.

الجدولان التاليان يوضحان العملية الحسابية المتعلقة بتوليد التوزيعات النسبية.

جدول (3-6): نوع المبحوثين

النوع (X)	التكرار (F)	النسب المئوية %
أنثى	8	$\frac{8}{20}\times 100 = 40$
ذكر	12	$\frac{12}{20}\times 100 = 60$
المجموع	**20**	**% 100**

جدول (3ـ7): الرضا الصحي

الرضا الصحي (X)	التكرار (F)	النسب المئوية %
غير راضٍ	7	$\frac{7}{20}\times100=35$
راضٍ	5	$\frac{5}{20}\times100=25$
راضٍ جداً	8	$\frac{8}{20}\times100=40$
المجموع	**20**	**% 100**

جدول (3ـ8): عمر المبحوثين

العمر بالسنوات (X)	التكرار (F)	النسب المئوية %
18	7	$\frac{7}{20}\times100=35$
19	5	$\frac{5}{20}\times100=25$
20	4	$\frac{4}{20}\times100=20$
21	2	$\frac{2}{20}\times100=10$
22	2	$\frac{2}{20}\times100=10$
المجموع	**20**	**% 100**

تجدر الملاحظة أن عمود النسب المئوية لابد أن يصل مجموعه إلى 100 %، لكننا في بعض الأحيان نجد أن النسبة في الجدول قد لا يصل مجموعها إلى 100 % وذلك نتيجة لعملية التقريب. فعلى سبيل المثال، قد تكون النسبة 22.3 %، 38.4 %، 39.3 %. فإذا تم تقريب هذه الأرقام إلى أقرب وحدة عشرية: 22 %، 38 %، 39 %، فإن هذه النسب المقربة يصل مجموعها إلى 99 %. وعندما يحدث ذلك، ينبغي على الباحث أن يبين في هامش تحت الجدول مشيراً بعبارة "النسبة لم تصل إلى 100، وذلك نتيجة لعملية التقريب" [7].

4- جداول التكرار المتجمع:

يمكننا القول بصفة عامة إن بناء جداول تكرارية لمتغيرات مقاسة على المستوى الترتيبي، وذي المسافات والنسبي إذا ما قورنت ببناء الجداول التكرارية البسيطة. ففي الجدول التكراري المتجمع تضاف أعمدة تتعلق بالتكرارات المتجمعة، والتكرارات المتجمعة النسبية. ولما كانت البيانات المقاسة على المستويين الترتيبي وذي المسافات والنسبي تسمح لنا بعملية ترتيب الحالات المدروسة من الأدنى إلى الأعلى، فإنه يفضل في هذه الحالة معرفة العدد و/أو النسبة للحالات التي تقع فوق أو تحت نقطة محددة من المقياس. فالجدول التكراري المتجمع يشير إلى كل قيمة في التوزيع وتجميع أعداد الفئات إلى بعضها البعض لتشمل تلك القيمة.

أما الجدول التكراري المتجمع النسبي Cumulative relative frequency table فهو يبين كل فئة في التوزيع، النسب المئوية والتناسب للعدد الكلي للحالات تجمع بعضها إلى بعض لتشمل تلك القيمة.

تجدر الملاحظة أن كل التكرارات المطلقة والنسبية والتكرارات المتجمعة لمتغير ما يمكن جمعها – في بعض الأحيان – في جدول واحد كما هو بيّن في الجدولين التاليين:

جدول (3ـ9) الرضا الصحي للمبحوثين

معدل الرضا الصحي	التكرار	التكرار المتجمع	النسبة المئوية %	النسبة المتجمعة %
غير راضٍ	7	7	35	$\frac{7}{20}\times 100 = 35$
راضٍ	5	12	25	$\frac{5}{20}\times 100 = 25$
راضٍ جداً	8	20	40	$\frac{8}{20}\times 100 = 40$
المجموع	**20**		**100**	**% 100**

جدول (3-10) عمر المبحوثين بالسنوات

العمر بالسنوات	التكرار	التكرار المتجمع	النسبة المئوية %	النسبة المتجمعة %
18	7	7	35	35
19	5	12	25	60
20	4	16	20	80
21	2	18	10	90
22	2	20	10	90
المجموع	**20**		**100**	**% 100**

وبتلخيص هذه التوزيعات بهذه الطريقة يمكن للباحث أن يجيب بشكل محدد على الأسئلة البحثية المطروحة. فإذا رغب الباحث في معرفة كم عدد المبحوثين الذين يبلغون من العمر 18 أو 19 سنة.

فالباحث ببساطة ينظر إلى مجموع الحالات في أول الصفين من الجدول رقم (10) ليجد أن التكرار المتجمع يصل إلى 12، أي 60% من كل الحالات. وبالطريقة نفسها، إذا رغب الباحث في معرفة كم عدد الحالات التي تزيد أعمارها عن 19 سنة، فإنه باستطاعته ملاحظة أن 60% من الحالات أعمارهم 19 سنة أو أقل، فإن ذلك يترك 40% من الحالات فوق هذا العمر (40% = 60 - 100)[8].

تجدر الإشارة إلى نقطة إضافية في هذا السياق هي أنه عند التعامل مع المتغيرات المقاسة على المستوى ذي المسافات والنسبي عادة ما تستخدم فئات متساوية Class Interval بدلاً من قيم فردية لبناء توزيع تكراري.

وعند بناء التوزيع التكراري للفئات المتساوية يتطلب الأمر اختزال البيانات إلى عدد من الفئات الموجزة بدرجة يمكن التعامل معها، إلا أن هذا الاختزال قد يؤدي إلى ضياع بعض المعلومات، وعليه ينبغي على الباحث تجنب اختزال البيانات إذا كانت البيانات أصلاً قليلة وعرضت بشكل بسيط وبقيم فردية.

وَيُسْتَخْدَم التوزيع التكراري للفئات المتساوية إذا كان مدى القيم كبيراً جداً يصعب معه عرض البيانات وتحليلها. ولتوضيح بناء جدول تكراري يحتوي على فئات متساوية نسوق المثال التالي للتدليل:

جدول (3 ـ 11)

يحتوي على فئات متساوية: الدخل الأسبوعي لعشرين مبحوثاً

الفئات (X)	التكرار (F)
0 – 99	2
100 – 199	0
200 – 299	1
300 – 399	3
400 – 499	9
500 – 599	3
600 +	2
المجموع	**20**

من هذا الجدول يمكن ملاحظة أن أعلى توزيع للحالات التي تقع داخل الفئة (499 – 400)، كما يلاحظ أيضاً توزيع الدرجات عبر الفئات المتساوية الأخرى. لاحظ أننا في هذا الجدول لن نستخدم القيم الفردية التي تظهر في التوزيع لتبين كل صف، وبدلاً من ذلك تم استخدام حدود فئة محددة. فالحدود الحقيقية هي الحدود العليا والدنيا للفئات المتساوية التي تحدد نطاقاتها. وبشكل عام، إن الفئات المتساوية لديها نفس النطاق (الحجم)، بالرغم من أنه عند نهاية الحد الأدنى والحد الأعلى لمدى البيانات وضعت بشكل غير محدد open - ended كما هو الحال في الفئة من 600 فأكثر. إِن التطابق الفعلي للفئات المتساوية يعتمد على موقف معين، خصوصاً كمية المعلومات المطلوبة. وعليه، فإنه كلما اتسعت الفئات المتساوية سهل قراءة التوزيع، غير أن أقل المعلومات تكون في المتناول، فعلى سبيل المثال، إذا

استخدمنا فئات متساوية، ولنقل بمدى 200 مثل (199 – 0) و (399 – 200... الخ). فإن هذه الفئات ستؤدي في المحصلة النهائية إلى ضياع معلومات كثيرة. فالحالات التي تختلف كثيراً فيما يتعلق بالمتغير المرغوب في دراسته، فالشخص الذي يكسب 400 د.ل، والشخص الذي يكسب 560 د.ل شهرياً يمكن النظر إليهما بشكل متساوٍ. وبشكل عام، عندما نقوم بجمع قيم في فئات متساوية فإننا نفقد معلومات حول التباين الذي تحتويه هذه البيانات، وبالتالي كلما كان نطاق المسافة واسعاً بين الفئات زاد معه بشكل كبير ضياع المعلومات.

وبشكل آخر، إذا كان لدينا حجمُ ضيِّقُ للفئات المتساوية في جدول، فإنه باستطاعتنا أن نكتشف تباين أكبر في البيانات التي نتعامل معها، ولكننا لا نستطيع أن نعرض البيانات بشكل بسيط يجعل منها طيِّعة ومقروءة. فعلى سبيل المثال، إذا استخدمنا فئات متساوية بطول فئة 50 لبيانات الدخل (99 – 50، 49 – 0 إلى آخره) فإن عدد الصفوف في الجدول لن تختزل البيانات في شكل يمكن قراءته بحسب ما نريد. وعند بناء جدول يحتوي على فئات متساوية، يتطلب منا أن نضمن أن تكون المسافات مانعة التبادل Matually Exclusive، وبالتالي فإن اختيارنا لِـ 100 كطول للفئة في الجدول رقم (11) حيث الفئات كالتالي: 99 – 0، 199 – 100، 299 – 200... الخ. لاحظ أن الحد الأعلى لكل فئة لا يتلامس مع الحد الأدنى للفئة التي تليها. ويتضح أن ثمة تفاوت بين 99 و 100، 199 و 200، 299 و 300، وهكذا...

من المفاهيم الأخرى التي يمكن التعامل معها في إطار الفئات المتساوية، هو مفهوم مركز الفئة (Mid. Point (M للفئات المتساوية. ونعني بمركز الفئة، ببساطة، مجموع الحدود الدنيا والحدود العليا مقسماً على 2.

$$Mid.Point\ (m) = \frac{\text{الحد الأعلى + الحد الأدنى}}{2}$$

فمركز الفئة للفئة 99 – 0 هو:

$$M = \frac{0+99}{2} = 49.5$$

والجدول التالي يبين الحدود الحقيقية لدخل عشرين مبحوثا:

الدخل	مركز الفئة (m)	التكرار
0 – 99	49.50	2
100 – 199	149.50	0
200 – 299	249.50	1
300 – 399	349.50	3
400 – 499	449.50	9
500 – 599	549.50	3
600 +	649.50	2
المجموع		**20**

إن بناء جدول يحتوي الحدود الحقيقية ومركز الفئة للفئات المتساوية سيكون مفيداً عند الحديث عن مقاييس النزعة المركزية التي سنتناوله في موضعه من هذا الكتاب.

من التقنيات الأخرى الشائعة في تبويب البيانات في شكل مجموعات يمكن قراءته هي العشيرات Deciles. وبدلاً من استخدام قيم المتغير لتبويب الفئات، فالعشيرات تستخدم نسب محددة لبناء جدول حوله. فمجموع الحالات ينبغي أن يرتب ويُقسم إلى 10 مجموعات متساوية الحجم. إن هذا الإجراء، إجراء شائع الاستخدام في تحليل البيانات المتعلقة بالدخل. فعلى سبيل المثال، يمكننا ترتيب كل العائلات في مجموع محدد طبقاً للدخل، من الأسر الأكثر فقراً إلى الأسر الأغنى. وبعد ذلك تقسم إلى 10 مجموعات متساوية على أن يحتوي العشير الأول، 10 % من العائلات الأكثر فقراً، ويتضمن العشير الثاني 10 % من الأسر الأخرى حتى نصل إلى العشير العاشر الذي يحتوي هو الآخر على 10 % من الأسر الأغنى. وبالنظر إلى النسبة الكلية للدخل التي يحتفظ بها كل عشير، فإنه بإمكاننا أن ندرك توزيع الدخل وطبيعة التغيرات التي حدثت [9].

جدول (3 ـ 12) الدخل الإجمالي وفقا للعشير (أستراليا 1989)

العشير	حصة الدخل الإجمالي 1989 %
الأول	1.7
الثاني	2.8
الثالث	3.9
الرابع	5.2
الخامس	6.8
السادس	8.6
السابع	10.7
الثامن	13.4
التاسع	17.3
الأعلى	29.4

المصدر: George Argyrous, op.cit, P.53

من خلال هذه البيانات الواردة أعلاه، يتضح لنا أن توزيع الدخل الإجمالي لم يؤثر بشكل متساوٍ عبر الأسر (طبقاً لهذا المقياس).

إجراءات توليد التكرارات باستخدام SPSS:

1- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Analyze ⇦ *Descriptive Statistics* ⇦ *Frequencies*

2- اختر المتغير أو المتغيرات لتوليد جدول تكراري بالنقر على هذه المتغيرات: (تقود هذه العملية إلى الحصول على جداول التكرار بشكل تلقائي من خلال لصق أكثر من متغير في صندوق Variable (s). فإذا أراد الباحث كل المتغيرات (في هذا المثال ثلاثة متغيرات) فعليه لصقها معاً.

3- انقر على ◄ بحيث يتم لصق المتغير أو المتغيرات المختارة في المساحة تحت المتغير أو المتغيرات Variable (s) وهي قائمة المتغيرات التي من خلالها يتولد جدول التكرار.

4- انقر على ok.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

Frequencies

Statistics

	Age in years	**Health rating**	**Sex of respondent**
N valid	20	20	20
missing	0	0	0

Frequencies Table

	Frequencies	**Percent**	**Valid percent**	**Cumulative Percent**
Valid 18	7	35.0	35.0	35.0
19	5	25.0	25.0	60.0
20	4	20.0	20.0	80.0
21	2	10.0	10.0	90.0
22	2	10.0	10.0	100.0
Total	20	100.0	100.0	

Health rating

Valid percent	Cumulative Percent	Percent	Frequencies	
35.0	35.0	35.0	7	Valid unhealthy
60.0	25.0	25.0	5	healthy
100.0	40.0	40.0	9	Very healthy
	100.0	100.0	20	Total

Sex of respondent

	Frequencies	Percent	Valid percent	Cumulative Percent
Valid Female	8	40.0	40.0	40.0
Male	12	60.0	60.0	100.0
Total	20	100.0	100.0	

أحياناً يصادف الباحث أن بعض المبحوثين لا يجيبون على بعض الأسئلة ومن هنا تكون لديه بعض المعلومات المفقودة فإذا افترضنا أن الحالة رقم 19 والحالة رقم 20 (ذكور) لم يحددا جنسهم. فالمخرجات في هذه الحالة يمكن بيانها من خلال الإجراء التالي:

1- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Frequencies ⇐ *Descriptive Statistics* ⇐ *Analyze*

أمر للنوع سوف يولد المخرجات التالية:

Frequencies

Statistics Sex of respondent

N valid	18
Missing	2

Sex of respondent

	Frequencies	Percent	Valid percent	Cumulative Percent
Valid Female	8	40	44.4	44.4
Male	10	50.0	55.6	100.0
TOTAL	18	90.0	100.0	
Missing	2			
TOTAL	20			

شكل رقم (3-1) مخرجات Spss للتكرارات

أسئلة للمراجعة:

1- بين الاختلاف بين النسب المئوية والتناسب؟

2- لماذا يكون التناسب دائماً أصغر من قيمة نظيره في النسب المئوية؟

3- حوّل التناسب التالي إلى نسب مئوية:

أ- 0.01 ب- 0.13 ج- 1.24 د- 0.0045

4- حوّل النسب المئوية التالية إلى تناسب:

أ- 12 % ب- 13.4 % ج- 167 % د- 3.5 %

5- من البيانات التالية:

80، 79، 90، 66، 99. 37، 78
81، 53. 78، 71، 57. 68، 89
60، 76، 98، 79. 93، 58. 77
69، 74، 80، 83، 92، 49. 77
75، 48، 73، 74، 84، 62، 90
59، 65، 87، 84، 75، 32، 98
72، 85، 62، 70، 55، 95، 86، 63

المطلوب: بناء جدول يحتوي على فئات متساوية بطول (10)

$$\begin{pmatrix} 30 & - & 39 \\ 40 & - & 49 \end{pmatrix}$$

وهكذا...

الهوامش والمصادر:

أولا: الهوامش:

1- George Argyrous, Statistics for Social & Health Research, with A Guide to Spss, SAGE Publications, London, 2001, P.38.

2- Ibid, P. 43.

3- Hubert M. Blalock. Jr. Social Statistics, Mac Graw Hill Book Company, INC, New york, 1972, P.34.

4- Ibid, P. 46.

5- Ibid, P. 46.

6- مصطفى خلف عبد الجواد، الإحصاء الاجتماعي: المبادئ والتطبيقات، دار المسيرة للنشر والتوزيع والطباعة، عمان، 2009، ص 40.

7- George Argyrous, Statistics for Social & Health Research, op. Cit, P.47.

8- Ibid, PP. 47 - 48.

9- Ibid, P. 51, 52 ,53.

ثانيا: المصادر:

1- George Argyrous, Statistics for Social & Health Research, with A Guide to Spss, SAGE Publications, London, 2001.

2- Hubert M. Blalock Social Statistics, Mac Graw Hill Book Company,INC, New york, 1972 .

3- Joseph F. Healey, The Essentials of Statistics A Tool for Social Research, Wadsworuh Cen Gage Learning, USA. 2010.

4- J. Richard Kendrick, Social Statistics: AN Introdaction using SPSS for Windows 2 ed, USA, New york, 2005.

5- مصطفى خلف عبد الجواد، الإحصاء الاجتماعي: المبادئ والتطبيقات، دار المسيرة للنشر والتوزيع والطباعة، عمان، 2009.

الفصل الرابع
مقاييس النزعة المركزية ومقاييس التشتت

أولاً: مقاييس النزعة المركزية:

تحتوي مقاييس النزعة المركزية على المنوال والوسيط والمتوسط وتساعد هذه المقاييس في إيجاد أنماط من البيانات كما أن لهذه المقاييس استخدامات في الحياة اليومية.

وتشير مقاييس النزعة المركزية إلى قيمة متوسطية في التوزيع وتتخذ هذه المقاييس إشارات مختلفة كما يتبين في الجدول التالي، وأن اختيار هذه المقاييس يمكن أن يتم من خلال حسابه من أي مجموعة من البيانات التي تتغير من خلال المستوى الذي تم على ضوئه قياس المتغير.

جدول (4-1) مقاييس النزعة المركزية

مستوى القياس	مقاييس النزعة المركزية
الاسمي	المنوال
الترتيبي	المنوال + الوسيط
ذو المسافات والنسبي	المنوال + الوسيط + المتوسط

المصدر: George Argyrous, Statistics for Social & Health Research, with a Guide to SPSS, SAGE Publications, London ,2001, p.64

نلاحظ من خلال هذا الجدول أن القواعد الأساسية في الإحصاء: هي تقنيات يمكن تطبيقها على مستوى معين من القياس، وكذلك يمكن تطبيقها على مستوى أعلى من القياس. فعلى سبيل المثال، إن مقياس النزعة المركزية يمكن حسابه من البيانات الاسمية (المنوال) وكذلك يمكن حسابه من البيانات الترتيبية والبيانات ذات المسافات المتساوية والنسبية وبطريقة مشابهة فإن المقاييس التي يمكن حسابها من البيانات الترتيبية يمكن حسابها أيضاً من البيانات ذات المسافات المتساوية والنسبية. وتجدر الإشارة هنا إلى أن المقاييس التي يمكن حسابها من مستوى معين من القياس ليس باستطاعتنا دائماً حسابها من مستويات أقل، فالمتوسط الحسابي على سبيل المثال، يمكن حسابه لمستوى أعلى من القياس (ذي المسافات والنسبي) ولتوضيح ذلك في كيفية حساب كل واحد من مقاييس النزعة يمكننا بيان ذلك من خلال البيانات النظرية التالية:

جدول (4-2) نوع المبحوثين

النوع	التكرار
ذكور	12
إناث	8
المجموع	20

جدول (4-3) الرضا الصحي للمبحوثين

الرضا الصحي	التكرار
غير راض	7
راض	5
راض جداً	8
المجموع	20

جدول (4-4) عمر المبحوثين

عمر المبحوثين بالسنوات	التكرار
18	7
19	5
20	4
21	2
22	2
المجموع	20

من خلال بيانات جدول (4 – 4) يمكننا حساب المنوال والوسيط والمتوسط، في حين يمكننا من الجدول (4 – 3) حساب المنوال والوسيط، أما في جدول (4 – 2) فإنه ليس باستطاعتنا سوى حساب المنوال.

المنوال:

دعنا نبدأ بمقياس المنوال (Mode) وهو أبسط مقاييس النزعة المركزية حيث يمكن حسابه على كل مستويات القياس.

ويعرف المنوال: بأنه القيمة الأعلى تكراراً في التوزيع أي القيم التي تتكرر أكثر من غيرها. والمنوال هو المقياس الوحيد للنزعة المركزية الذي يمكن حسابه من البيانات الاسمية والترتيبية وذات المسافات المتساوية والنسبي، وتعتبر هذه ميزة على الخيارات الأخرى (الوسيط والمتوسط) وذلك لسهولة حسابه. فإذا نظرنا إلى التوزيع التكراري لعشرين حالة حسب النوع (ذكور 12 حالة) و (إناث 8 حالات) فإن أعلى تكرار لهذه البيانات هم فئة الذكور، ومن هنا فإن التوزيع الأعلى تكراراً هو توزيع الذكور (Mo).

تجدر الإشارة إلى أنه عند التعامل مع المنوال يجب أن نضع في أذهاننا نقطة أساسية وهي أن المنوال هو قيمة المتغير الأكثر تكراراً وليس عدد الأوقات التي تظهر فيها القيمة في التوزيع[1].

ثمة خاصية ترتبط بالمنوال قد لا تنطبق على الوسيط أو المتوسط كمقياسين من مقاييس النزعة المركزية باعتبار أن التوزيع المرتبط بالمنوال قد يحتوي على أكثر من منوال، فعلى سبيل المثال، نفترض أن لدينا التوزيع التالي المتعلق بمتغير العمر كما هو موضح في الجدول (5).

على تصريحات شفهية عشوائية، ينقلها إعلاميون غير متخصصون فيتم عرضها بشكل قد يخالف الوقائع والضوابط العلمية بشكل تام.. ولا ينسى أحد أبدًا العبارة التي تقال دائمًا حول أية مواضيع يستهان بمصداقيتها «كلام جرائد» !

ولكتابة المراجع داخل المتن وكقائمة نهائية طريقة محددة يفضل تأجيل عرضها إلى الجانب التفصيلي في وقت لاحق.

تاسعًا: قائمة المحتويات Contents

وهي تمثل الخريطة التفصيلية المرحلية المتتابعة لمكونات البحث بشكل تصنيفي واضح وعادة ما تبدأ بالعموم وتنتهي بدقة الخصوص، وعادة ما يكون الحد الفاصل الدال على الانتقال إلى جزئية جديدة هي كلمة «المقدمة Introduction».

وعادة ما يتم إدراج المحتويات وفقًا للشكل التالي الأكثر شيوعًا.

مقدمة

(1) الباب

مقدمة

(1) الفصل [1/1]

مقدمة

(1) الجزء [1/1/1]

(2) الباب

مقدمة

(1) الفصل [1/2]

مقدمة

(1) الجزء [1/1/2]

ويقصد بالأرقام التصنيفية السابق الإشارة إليها الإشارة أولاً إلى رقم الباب (الأوسع مجالاً) ثم يتبع برقم الفصل محل العرض (الأكثر تخصصًا) ثم يتبع برقم

القسم أو الجزء (الأكثر تحديدًا) داخل الفصل المذكور، ويمكن أيضًا أن يتبع ذلك بفرعيه داخل القسم أو الجزء بإعطائها رقمًا تصنيفيًا رابعًا فيقال مثلا [1/1/1/1] حيث يقصد به (بدءًا باليسار بأكثر الجزئيات تخصيصًا): الفرعية الأولى من الجزء الأول في الفصل الأول التابع للباب الأول.

ملاحق

قائمة المراجع

وعادة ما تكون قائمة المحتويات على باب يتناول الأمور بعمومها أو ملخصات تقييمية لما ورد بالدراسات الأخرى ثم يتبع ذلك باب تعريفي بالموضوع محل الدراسة والتحليل بالاستعانة بالقواميس اللغوية والمتخصصة وبالدراسات ذات الصلة، ثم في باب ثالث يتم البدء في عرض وصفي وتصنيفي لكيفية الحصول على البيانات ولوسائل قياسها وإعدادها للتحليل، وفي باب رابع يتم إجراء التحليل العملي باستخدام أدوات التحليل المناسبة وتعرض من خلاله نتائج التحليل بشكلها الأولى البحث (مثلاً أن معامل الارتباط وجد أنه 0.3 وأن R^2 كانت 1، وذلك بدون أي توضيح لمعنى تلك النتائج) وفي باب خامس يبدأ الباحث في لعبته الفكرية الكبرى وإظهار مواطن تميزه الفكري الناقد والتقييمي من خلال تقديم تعقيباته وملاحظاته على ما ورد من نتائج التطبيق العملي مع ربط ذلك بما ورد بالدراسات السابقة أو دعمه بما اشتملت عليه من تفسيرات - ثم يتبع ذلك بخلاصة واستنتاج ومقترحات يتم تقديمها إلى من يهمه الأمر مع التذييل بالملاحق اللازمة ثم بقائمة المراجع على نحو ما بيناه وما سوف تقوم بشرحه بتفصيل أكثر عند عرض كيفية الكتابة النهائية للبحث العلمي.

الفصل الثاني

تحديد البيانات ومصادرها
وكيفية تحصيلها وأدوات تحليلها

مقدمة:

بالرجوع إلى مثالنا اللذيذ حول كيكة الزبادي – عرفنا أن ربة المنزل قد انتهت من تحديد نوع الوصفة التي تفضل اتباعها من أجل إعداد الكيكة؛ ومن ثم كان عليها أن تتعرف على الجهات التي سوف تحصل منها على مستلزمات إعداد الكيكة فالقالب الخاص يشترى من كل الأدوات المنزلية، والدقيق والسكر والزبدة والبيض والخميرة يتم شرائها من محل بقالة موثوق في أن مبيعاته صحية وغير منتهية الصلاحية؛ وهكذا..

وعلى ذات النهج تمامًا، يجب على الباحث أن ينحو من أجل الحصول على البيانات والمعلومات الصحيحة والملائمة بطبيعة البحث ولاحتياجاته بحيث يتمكن من استخدامها لإتمام العملية الوصفية والتحليلية والتقييمية اللازمة لإنجاز بحثه على أكمل وجه ممكن.

وبناء عليه، يتم عرض ذلك الفصل كبيان إرشادي من أجل توعية الباحث بكيفية إنجاز عملية صحيحة متكاملة لتحصيل البيانات الملائمة من أفضل مصادر ممكنة لها. وسوف يتم ذلك العرض من خلال ثلاثة أجزاء يتناول أولها كيفية التحديد الأفضل للبيانات اللازم توافرها لإنجاز البحث المستهدف، ويستطرد في الثاني توضيح وسرد بعض المصادر الهامة التي يمكن للباحث أن يعثر على البيانات المستهدفة في أحدها أو بعضها أو كلها، ويعرض في ثالثها بعض الأساليب والطرق الشائع استخدامها من أجل إنجاز عملية تحصيل ناجح للبيانات بمختلف أنماطها ودرجات تيسر أو صعوبة الاستحواذ عليها.

1/2/2 كيفية تحديد البيانات اللازمة لإنجاز البحث Data specification

يتم ذلك من خلال عملية تكاملية تتطلب الآتي:

أ - الاسترشاد بالدراسات المماثلة والتعرف على طبيعة كل دراسة وعلى البيانات التي قامت باستخدامها وعلى مدى نجاح تلك المفاضلة والخيارات وفقًا لما تتم ملاحظته من نتائج وتعقيبات ومقترحات. وذلك ما سوف يتم توضيحه بمزيد من الانضباط من خلال الجزء المدون حول الدراسات السابقة كما سوف يتضح في حينه.

ب- إجراء عملية بحثية أولية لمصادر البيانات المحتملة، والتعرف على مدى إمكانيات استجابة تلك المصادر لتوفير البيانات المرجوة.

ج- الاستعانة بخبرات الباحثين الآخرين وبخبرات أهل العلم والخبرة خاصة في المجالات الإحصائية والمعلوماتية المتخصصة.

د - التأكد من ملاءمة كل مصدر للإمكانيات المتاحة للباحث من لياقة بدينة، وحالة مادية تمويلية، .. الخ بالإضافة إلى تفضيل النوعية الخالية من متطلبات الإجراءات البيروقراطية المعرقلة.

هـ - وضع تصور عام لمفردات الجزء الوصفي والتحليلي اللازم لتحقيق هدف البحث،

وتحديد تصور محدد لمتغيرات المفاتيح الصورية dummy والرقمية Numerical التي يزمع استخدامها بالبحث وبالوصف وبالتحليل، ثم تحديد البيانات التي يجب أن تتوافر من أجل قياس تلك المتغيرات.

فمثلاً إن كان الباحث يريد إجراء بحث حول سوء التغذية في بلد ما، وعرف من خلال الدراسات الطبية المتخصصة أن سوء التغذية يتم قياسه بمقاييس مختلفة مثل مقياس جوميز Gowez، ومقياس واترلو Waterlow، فليركز انتباهه على كل مقياس منهما وليبحث على المفردات التي يقوم عليها، فإن كان الأول يقوم على قياسات ترتبط بمتغير العمر age والوزن weight، بينما المقياس الثاني يركز على متغيرات الطول height والوزن weight. فعلى الباحث أن يبحث عن أكثر البيانات يسرًا في الحصول عليها لكي يطبق أحد هذين المقياسين.

فإذا فضل - مثلاً- معيار جوينز، تكون البيانات المطلوبة في تلك الحالة هي حول الأعمار المختلفة لفئة معينة، والأوزان المقابلة كلٍ وفقًا لعمره.

فإذا كان الباحث متخصصًا فعليه أن يستخدم الوسائل الطبية العملية الحرفية لاستجلاب البيانات المطلوبة من خلال عينة من زوار المراكز الغذائية مثلاً. أما إذا كان الباحث غير متخصص لكنه يود قياس ذلك بنفسه من أجل تطبيقه على دراسة متخصصة ذات طبيعة اجتماعية أو اقتصادية مختلفة (مثل دراسة كيفية تأثر تلك الحالة الغذائية بالحالة الاقتصادية والاجتماعية للأمهات)، فعلى الباحث غير المتخصص أن يبحث عن بيانات أكثر ملاءمة ليس فقط لتخصصه غير المختلف، ولكن أيضًا للمتغيرات الأخرى غير الطبية التي تتطلب قياسات مختلفة وبالتالي بيانات أخرى قياسية مختلفة مثل الحالة التعليمية للأمهات ومتوسطات ما يحصلن عليه من دخول ... الخ[1].

(1) لمزيد من التفاصيل أرجع إلى: زينب صالح الأشوح، 1994، ص ص 43-50.

ويوجه عام، تجب ملاحظة أن كلمة البيانات Data تشير إلى المواد الخام raw التي ستستخدم من خلال بعض المعالجات العلمية المناسبة للخروج بنتائج تتلاءم والأهداف المحددة مسبقًا للبحث،ومن ثم فمثل تلك البيانات يمكن أن تتمثل في عناصر خام أصلية، أو معلومات ونتائج نهائية في دراسات أخرى لكن تم النظر إليها كبيانات أولية تتطلب معالجة بحثية تحليلية جديدة في البحث الجديد. وقد تكون تلك البيانات ذات طابع أو عرض إحصائي رقمي أو بيانات لفظية خليط من هذا وذاك.

ولكن يا تري! من أين يمكن أن يتم تحصيل البيانات المستهدفة على النحو المأمول ذلك ما نود توضيحه من خلال الجزء التالي.

2/2/2 مصادر البيانات Data resources

الواقع أن مصادر البيانات يمكن أن تصنف وفقًا للمنظور المصوب إليها. وبوجه عام، يمكن تحديد مصادر البيانات التي يجب على الباحث أن يعرف منها على ما يحتاجه من خلال التصنيفات التالية.

أولا: من حيث درجة المصداقية والتوثيق

(1) مصادر أولية Primary Data:

وهي المنابع الأساسية التي تم تدوين وعرض البيانات فيها، حيث تعتبر البيانات المأخوذة منها بمثابة بيانات «أول قطفة» وهي تتميز بالمصداقية الأكبر وبالتالي يكون تحليليها واستخراج المعلومات من خلال تحليلها ومعالجتها هو الأكثر صوابًا ومصداقية وإقناعًا ومنطقية.

ومن الأمثلة على البيانات المستقاة من مصادرها الأولية، الإحصائيات السكانية والإسكانية الواردة في النشرات الرسمية الدورية الخاصة بها، والبيانات الداخلية عن العاملين داخل المؤسسات المختلفة والواردة في السجلات الرسمية التابعة لها،

والبيانات التي يصرح بها الوحدات محل الدراسة من خلال الدراسات الميدانية والمقابلات الشخصية المباشرة والبيانات التجريبية أو المعملية المنظورة أو الملموسة ... الخ.

وعلى وجه العموم يمكن أن تحدد أهم المصادر الأولية للبيانات على اختلاف أنواعها فيما يلي:

أ - الكتب والكتيبات والتقارير والدوريات الشهرية أو السنوية المتخصصة على المستوى المحلى أو العالمي (وتضم عادة بيانات وأرقام وحقائق عن الأنشطة الخدمية والإنتاجية والسياسية والاجتماعية وغيرها مما يخص الدولة أو المؤسسات المحلية أو الإقليمية أو الدولية) وذلك مثل الكتاب السنوي للجهاز المركزي للتعبئة العامة والإحصاء والكتاب السنوي للأمم المتحدة، والتقرير السنوي للبنك المركزي ومجلات البحوث العلمية، وتقارير التنمية البشرية المحلية والدولية .. إلخ.

ب- المطبوعات الرسمية الحكومية (مثل دليل النظام في كل جهة حكومية وسجلات العمالة وسجلات المرتبات والأجور، وتقارير الأداء والإنتاجية).

ج- وثائق رسمية جارية: مثل مخاطبات ومراسلات رسمية وعلمية تدور بين المؤسسات المعنية المختلفة وتشتمل على معلومات حول أنشطتها المختلفة.

د - الأطالس: مراجع جغرافية حول الموارد الطبيعية والمواقع الجغرافية وطبيعة الطقس المرتبطة بالدول والبحار والقارات وما شابه ذلك.

هـ - المعاجم والقواميس اللغوية والمتخصصة: وتفيد في تحصيل البيانات اللغوية المختلفة وتلك المتعلقة بالمفاهيم وبالمصطلحات اللغوية الهجائية والعلمية المتخصصة مثل القواميس الطبية والقواميس الهندسية والاقتصادية، ومعاجم اللغة العربية مثل قاموس المحيط.

و - براءات الاختراع: وهي الوثائق التي تسجل اختراع شيء جديد لم يكن معروفًا ولم ينشر حوله أو عنه خبر سابق.

ز - التراجم والسير الذاتية خاصة فيما يتعلق بكبار الشخصيات العلمية أو القيادية أو السياسية.

ح- وثائق تاريخية محفوظة ومخطوطات تم تدوينها بواسطة أشخاص موثوق فيهم مثل تلك التي تمثل التراث العربي والإسلامي.

(2) مصادر ثانوية Secondary Data:

وهي عادة ما تكون منقولة من المصادر الأولية أو مبنية على أساسها ومن ثم فهي تعتبر أقل دقة ومصداقية نتيجة لزيادة احتمالات الخطأ في متضمناتها من بيانات كنتيجة لزيادة احتمالات النقل عن الأصل، أو التعبير المستحدث أو الترجمة أو سوء الفهم الفعلي غير المقصود أو للتحريف المتعمد لأغراض التشويه أو التضليل. ومن أنواع تلك المصادر الثانوية:

أ - الكتب والدوريات المتقطعة أو غير المنتظمة (وتسمى مطبوعات مسلسلة).

ب- الموسوعات ودوائر المعارف: حيث تشتمل على خليط من المعلومات المشتقة من مصادر أولية أو من مصادر ثانوية أخرى.

ج- الكتيبات والنشرات غير الرسمية والتي تصدر لأغراض غير علمية مثل الدعاية والإعلان والانتخابات.

د - الأبحاث والمجلات غير المحكمة من قبل لجان علمية معتمدة.

هـ - الوسائل الإعلامية خاصة غير المتخصص منها.

ثانيا: من حيث وسيلة وطبيعة العرض:

1- مصادر ورقية مثل المطبوعات المختلفة من كتب ورسائل علمية ومجلات متخصصة التي تتوافر عادة في المكتبات العامة والخاصة ودور النشر وما ماثلها.

2- مصادر بصرية كالمشاهدات والظواهر التي تدرك بالبصر والتبصر وبالملاحظة، وذلك فيما يتعلق بشكل مباشر أو غير مباشر بوضع البحث سواء في أماكن مخصصة مثل المعامل أو في ميادين تواجد حالات الدراسة أو ما ماثل ذلك.

3- مصادر سمعية مثل ما يستجلب من خلال المقابلات الشخصية أو المؤتمرات أو الندوات أو الحلقات النقاشية أو المحاضرات العامة والمتخصصة ... الخ.

4- مصادر إلكترونية: وهي تلك التي تتيحها تكنولوجيا المعلومات من خلال تحويل المجموعات الورقية إلى أشكال جديدة إلكترونية سهلة العرض أو الاستجلاب والاستخدام وذلك مثل شبكة الإنترنت واستخدام الوسائل الممغنطة والأقراص الليزرية المكتنزة (DVD) ... الخ.

وفي جميع الأحوال فإن هناك ملحوظة بالغة الأهمية عند تحديد المصادر المزمع استخدامها لاستجلاب البيانات المستهدفة، وهي ضرورة إعطاء الأولوية المطلقة للمصادر الأولية وعدم استخدام المصادر الثانوية إلا في أضيق الحدود وفي حالات الضرورة القصوى حين لا يتوافر المطلوب في المصادر الأولية. وفي تلك الحالة الأخيرة، لابد أن ينظر إلى نتائج التحليل للبيانات الثانوية بعين التشكك والريبة تحسبًا لاحتمالات الخطأ مثلمًا أشرنا آنفًا.

ويتبقى لنا إذن أن نعرض لباحثينا الكرام، أهم الطرق والوسائل التي يمكنه الاستعانة بها لكي تتم عملية استجلاب وتحصيل البيانات بأكبر كفاءة ممكنة حيث تقوم تلك الكفاءة على عناصر رئيسية هي **كفاية حجم البيانات**؛ و **ملاءمة نوع البيانات** للعمليات التحليلية المزمع إجرائها، وللأدوات القياسية التي سوف تستخدم في معالجتها، ولتخصص الباحث وطبيعة البحث؛ وكذلك من مقومات كفاءة البيانات محل التحصيل والتحليل أن تكون **على أكبر قدر من الدقة والمصداقية** وأن يتم الحصول عليها **في حدود الوقت المتاح** و **القدرات والإمكانات المتاحة** ويكون كل ذلك **بأقل تكلفة ممكنة**.

3/2/2 أهم وسائل وأدوات وطرق تحصيل البيانات واحتوائها

Data Collection, Containers and Techniques

قامت ربة المنزل بالمفاضلة بين الوصفات المختلفة لكيكة الزبادي من خلال استقرائها المطبوعات المختلفة ومشاهدتها لبرامج الطهو المختلفة أو سماعها لها من خلال وسائل الإعلام المختلفة أو من بعض خبراء الطهو من صديقات وغيرهن.. كما عرفت من خلال تحرياتها من أين تأتي بمستلزمات الكيكة أصبح عليها إذن أن تذهب إلى أماكن تواجد عروض وتسويق تلك المستلزمات من أجل استجلاب اللازم منها.. ولكن لأن بعض المستلزمات تكون سائلة، وبعضها ثقيلة، وبعضها سهلة التلف، كان عليها أن تأخذ معها أنواعًا مختلفة من الحاويات لتحتفظ فيها بالمستلزمات حتى رجوعها إلى منزلها سالمة بها وحتى يتأتى لها استخدامها في إعداد الكيكة بسلاسة ويسر. وبعد عودتها إلى المنزل ذهبت إلى (المطبخ) وقامت بتفريغ المستلزمات ووضعها في أواني وحاويات مناسبة كلٌ لكل، وبشكل منفصل وواضح ثم قالت بسم الله الرحمن الرحيم، فلأبدأ في إعداد كيكتي اللذيذة..

هذا تمامًا هو ما يجب على الباحث الماهر أن يتبعه لكي يعد بحثه في مطبخه التحليلي العلمي المتخصص. فقد انتهى من تحديد البيانات اللازمة لإعداد بحثه وفقًا للأهداف المحددة سلفًا للمناهج العلمية التي وجدها الأفضل لإنتاجها وذلك من خلال الاسترشاد بالخبراء والمختصين من خلال الدراسات المتخصصة أو المقابلة مع الخبراء أو ما شابه ذلك. وعلى نفس النهج، كون حصيلة معرفية واسعة وواضحة ودقيقة حول المصادر التي تتوافر فيها مثل تلك البيانات التي يمكن أن تتمثل في إحصائيات مطبوعة أو مرتقبة الاستجلاب من ألسنة المختصين، ويمكن أن تكون مواد سائلة أو صلبة أو حيوانات تجارب أو جثث تعليمية... الخ. على الباحث إذن أن يذهب بنفسه لتحصيل المقرر من مثل تلك البيانات، وأن يعد الحاويات المناسبة التي يجب استخدامها من أجل الاحتفاظ بالبيانات المستجلبة لحين البدء بعمليات استخدمها على النحو المستهدف.. كلٌ في أوانه وفي حينه، هيا إذن

لنتعرف معًا على نماذج لبعض تلك الوسائل والحاويات كما قررها المختصون في مجال البحث العلمي.

أولا: أهم الأدوات والحاويات اللازمة لعمليات تحصيل البيانات وحفظها

الواقع أننا نفضل دائمًا استخدام صور من حياتنا المعيشية الدارجة من أجل تقريب واقع ما نقوم بدراسته وبتحليله ومن أجل الاستفادة بكل الخبرات حتى غير المباشر لتخصصنا مما يضمن إتمام بحثنا على أفضل حال علمي معرفي ممكن. ومن ثم، فإنه بالإضافة إلى مثال طاهية الكيكة، يجب أيضًا أن نضيف نموذجًا تصوريًا أكثر جدية والتصاقًا بمهام الدارس الديناميكية الباحث دومًا عن الحقيقة بكل مكامنها وخباياها ودقائق مكوناتها. ذلك التصور الآخر ذكره الدكتور أحمد بدر في كتاب له حول أصول البحث العلمي ومناهجه[1]، حيث صور عملية البحث والتقصي عن البيانات بالنسبة لأي دارس جاد، كعملية البحث والتقصي والتحري من ملابسات جريمة ما والتي يجب أن يقوم بها فئات مدربون بحنكة على ذلك من العاملين بأجهزة الشرطة، وبالمحاكم، ومن المحامين وما هم على شاكلتهم. وأوضح المؤلف أن ما يقومون بالبحث عنه يكون بمثابة أدلة دامغة تكشف حقائق وملابسات الجريمة ومن ثم فلابد أن يكون ما حصلوا عليه من أدلة (وهي تقابل البيانات في مجالنا الحالي) لابد أن تكون **مادية** ومستمدة من مصادر صحيحة وموثوق بها.

وبالاستعانة بتلك الصور التمثيلية المرشدة، وبالرجوع إلى الدراسات المتخصصة، وبدمج ذلك مع خبرتنا الكثيفة طويلة الأجل في مجال البحث العلمي تعلمًا وتعليمًا وإشرافًا وتقييمًا وممارسة، يمكن القول بأن أهم الأدوات الواجب استخدامها من أجل استجلاب البيانات وتحصيلها بأفضل طريقة ممكنة، يمكن أن تتمثل فيما يلي - على سبيل المثال، لا الحصر:

---

(1) مرجع سبق ذكره، ص 55.

1- **الاستبيان Questionnaire**: صحيفة الاستبيان والاستقصاء والاستخبار.

2- **الاستبار Interview**: صحيفة المقابلة الشخصية.

3- **الملاحظة Observation**: مدونة تصنيفية حول المشاهدات وانطباعات الباحث الفورية وتعقيباته عليها.

4- **الاتصالات Communications**: من خلال الهواتف الأرضية والمحمولة وشبكات التواصل الاجتماعي والبريد الإلكتروني.

5- **المراسلات E-mail or Mails**: وعادة ما تكون مراسلات معدة بطريقة رسمية وواضحة بشكل يماثل استمارات الاستقصاء لكن يفضل اختصارها تجنبًا للملل المتلقي وانصرافه عن الرد.

6- **الإحصاء والتعداد Census**: بحيث يقوم به الباحث مباشرة أو ينيب عنه متخصصون كأفراد أو كفرق.

7- **المقاييس Measurements**: ويفضل أن يقوم الباحث مباشرة أو يستعين بمتخصصين بشكل جزئي وفي حالات الضرورة فقط.

8- **التصوير الساكن أو المتحرك، الصامت أو بالصورة والصوت Photocopying, Photochronographing**: وذلك باستخدام الأجهزة المناسبة من كاميرات وأجهزة DVD وغيرها.. حيث تصلح تلك الأداة خاصة للقائمين بالدراسات العملية التطبيقية مثل الجيولوجي والمهندس المعماري والصيدلاني والمهندس الزراعي والباحث الاجتماعي لحالات في مواقع معينة مثل المنحرفين داخل السجون ... الخ.

9- **المعدات الفنية Technique Tools**: مثل المعدات اللازمة لوزن الفئة المستهدف دراستها ولقياس أطوالهم في دراسات حول الحالة الغذائية لهم، أو الأدوات الدقيقة لاستجلاب عينات من نبات معين.. الخ.

10- **الرسوم البيانية والأشكال الهندسية والخرائط التوضيحية للمواقع المستهدفة**

Graphics, Maps ...: وذلك من أجل إرشاد الباحث إلى المواضع والحالات الصحيحة المستهدف جمع البيانات منها أو حولها أو بشأنها.

11- **الوسائل الممغنطة وغيرها من وسائل استجلاب المصادر الإلكترونية**: مثل الاسطوانات والفلاشات[1].

وبالإضافة إلى ما سبق، فهناك أدوات أخرى ترتبط بالبيانات المشتقة أو المحصلة من المنشورات والمؤلفات والرسائل العلمية والأبحاث المتخصصة، ومن أهم تلك الأدوات كروت أو قصاصات ورق أو مذكرات أو كشاكيل متنوعة الأحجام والأطوال، وأقلام ملونة، فلوماستر ورصاص وجاف وفوسفورية..، وكوريكتور وغيره من أنواع الإزالة السريعة لغير المرغوب فيه مما تم تدوينه.

وعادة ما تستخدم مثل تلك الأدوات أيضًا بشكل تكميلي بالإضافة إلى ما سبق عرضه في حالات تجميع البيانات من المصادر الأخرى على اختلاف أنواعها وأنماطها ومصادرها.

وتجنبًا للإطالة الباعثة على الملل والمسببة للتيه وللخلط، نود الاكتفاء بهذا القدر من الأدوات والوسائل، تاركين لكل باحث فطن مهمة الإضافات الأخرى التي تدعمه في عملية استجلابه للبيانات وللمعلومات بأكثر درجات الإتقان، وبأكثر درجات الكفاية الممكنة، وبأقل التكاليف والمهدرات والخسائر الممكنة.

لكن.. إن كنا توقفنا هنا عن الحديث بالمزيد حول وسائل وأدوات تحصيل البيانات وحفظها، فعلينا ألا نغفل عن شرح كيفية تفعيل تلك الوسائل والأدوات وتوظيفها بالشكل الصحيح من أجل تحقيق الدور الذي تمت الاستعانة بها من أجله وهو ما نزمع عرضه توا في الجزئية التالية.

---

(1) بعض ما سبق عرضه تم اقتباسه من بعض المراجع، وهي: محمد شفيق، 2005، مرجع سباق ذكره. وزينب الأشوح، مرجع سبق ذكره.

ثانيًا: كيفية تحصيل البيانات وتصنيفها وإعدادها بالشكل الصالح لاستخدامها في التطبيقات العملية للبحث المستهدف إنجازه

وسوف يتم التركيز هنا بتوضيح تفصيلي متعمق وشامل على استمارة الاستبيان Questionnaire حيث أنها من أكثر الوسائل التي تتطلب إعدادها بشكل دقيق وقائم على ضوابط علمية محسوبة ومحددة وبطريقة ملزمة. خاصة وقد لوحظ أن الكثيرين يخطئون كثيرًا في إعدادها اعتقادًا منهم بأنها خطوة ترتبط بجهدهم البحثي الشخصي، وبأن ما للمحكم أو للقارئ عندهم هو فقط الثمرات النهائية لتحليل ما تضمنته من بيانات، مع أن كثير من الباحثين يفشلون في تحصيل البيانات الصحيحة نتيجة لإخفاقهم في إعداد استمارة التحصيل بالبنود وبالمكونات وبالوسائل الصحيحة، مما يتسبب في النهاية إلى عرض نتائج مضللة، فما بني على خطأ، فهو خطأ.

ومع هذا، فسوف نقدم في نهاية العرض، مجموعة من الإرشادات لكيفية تفعيل استخدام بعض الأدوات الأخرى التي تتطلب ذلك مثل المقابلة الشخصية والكروت.

- كيفية إعداد واستخدام استمارة الاستقصاء وكيفية الإعداد الصحيح لدراسة ميدانية متكاملة

عادة ما يتم إجراء أية دراسة ميدانية متكاملة من خلال أربع مراحل أساسية هي:

1- القيام بدراسة استطلاعية أو إرشادية Pilot study.

2- تحديد أسلوب تصميم واختيار العينة Sample design.

3- تحديد كيفية جمع البيانات المطلوبة وكيفية استخدام استمارة الاستقصاء Questionnaire.

4- عرض المحددات والتحفظات التي يمكن أن تؤخذ على البيانات المحصلة من خلال الدراسات Considerations & Observations & Limitations.

وفيما يتعلق **بالمرحلة الأولى** المرتبطة بعملية **الدراسة الاسترشادية** Administration يجب البدء بإدراك وبإبراز أهميتها ولزوم إجرائها، ومن أهم ما يجب قوله في ذلك الصدد هو أنها تفيد في تحقيق هدف جوهري يتمثل في التحقق من صحة وفعالية استمارة الاستبيان التي يزمع استخدامها من أجل تسجيل البيانات المطلوبة، والتحقق من قابليتها للتطبيق العملي ومن تضمنها للبنود وللأسئلة التي يمكن بواسطتها الحصول على البيانات المستهدفة في أفضل شكل لها، وبحيث تمثل المجتمع محل الدراسة بأدق صورة ممكنة.

وعادة ما يتم استخدام مسودة توضيحية تكتب فيها جميع البيانات المطلوب تجميعها، مع عرض تساؤلات بديلة ترتبط الإجابات المستهدفة عليها بالبيانات المستهدف تحصيلها بشكل مباشر أو غير مباشر. وفيما يتعلق بالأمور الحساسة أو بالمواضيع التي يعزف الآخرون عن الحديث عنها أو عن التصريح بها مثل الأمور الشخصية أو التي ترتبط بالدخول وبالمكاسب المتحصلة أو بالتقييم الشخصي الذاتي يفضل أن توجه أسئلة أخرى داعمة وبشكل غير مباشر وفي بنود منفصلة متباعدة بحيث لا يفطن المرء المستهدف البيان منه إلى حقيقة الأمر ما يدفعه إلى رفض الإجابة أو إلى العزوف عن إكمال ضخ المطلوب من البيانات.

ويجب على الباحث أن يعرض بدقة السبب في كل سؤال، والهدف المتوقع من توجيهه، والصعوبات المتوقعة من صياغته أو من تقبل الآخر له، وكيفية مقابلة تلك الصعوبات بأسلوب علمي إيجابي فاعل. ولابد أن يحدد الباحث الثغرات والتحفظات المتوقعة وكيفية معالجتها. ويظل الباحث في استكمال الوضع المبدئي للاستمارة في تلك المسودة مع التصور الدائم لجدوى كل ما يفعله ولارتباطه اللصيق بأهداف البحث وبنطاقه ومناهج البحث ومحتوياته وبكل ما تمت الإشارة إليه في الخطة من عناصر أساسية.

وبعد ذلك يقوم الباحث بإعداد شكل نهائي لاستمارة الاستقصاء الأولية أو الاستطلاعية Preamble مع إرفاق ورقة بالتحفظات المذكورة وبدائل مواجهتها

المقترحة حتى يتم المفاضلة في اختيار الأفضل منها على ضوء التنفيذ العملي لتلك الاستمارة. هذا، وعلى الباحث أن يقرر كيفية ملء بيانات تلك الاستمارة ومن سيقوم بتدوينها.. هل ستكون بالبريد العادي أو الإلكتروني بحيث تصبح الاستمارة المعدة ملائمة لشكل الاستبيان البريد؟ Mailed Questionnaire أم أنها سوف تسلم للمصادر لكي يملئونها ويسلمونها حال الانتهاء من ذلك بمعرفتهم؟ أم سيقوم الباحث نفسه بملئها وبتدوين الإجابات بنفسه من خلال مقابلات مباشرة مع المصادر وقراءة كل سؤال عليهم بنفسه ثم تسجيل ما يمليه المصدر عليه؟ أو سيحدث ذلك من خلال الهواتف الأرضية أو الجوالة أو من خلال البريد الإلكتروني؟

ومن تجاربي الخاصة وقت مزاولتي لمثل تلك الأنشطة الميدانية، اتضح بأن الباحث يقوم بكل الخطوات اللازمة من الألف إلى الياء فيما عدا التصريحات بالإجابات، فتلك تترك للمصادر الحرية المطلقة لهم بذكرها بدون أية محاولات للتأثير المتحيز من قبل الباحث ولا حتى بإدعاء القيام بالإيضاح لما يغمض على المصدر. فمن دواعي الأمانة المطلقة أن يقوم الباحث بتلاوة ما كتب في الاستمارة كما هو وأن يعزف تمامًا عن أية محاولات مقصودة أو تلقائية من قبل المصادر لكي تكون هناك أية إضافات توضيحية من قبل الباحث لأنها ستصبح إضافات غير علمية وخارجه عن النموذج والإطار العلمي للاستمارة المعلنة، وذلك لاشك سيؤثر في مصداقية وفي دقة البيانات التي يتم تحصيلها.

ومن المنافع المكتسبة من تلك الاستمارة الاستطلاعية التعرف على مدى فهم المصادر للأسئلة الموجهة إليهم ومدى استيعابهم لمضمونها ومدى تقبلهم لكل محتوياتها واستعدادهم لتقديم إجابات كاملة على كل ما يوجه إليهم من أسئلة بدون ملل أو تخاذل أو مواربة.

بل، ومن خلال تلك الاستمارة الاستطلاعية يمكن أن تستوحى بنودًا أو أسئلة إضافية أو بيانات لم تكن في الحسبان، وذلك من خلال التعامل مع المصادر. في عقر أماكن تواجدها واستقرارها.

وبوجه عام، يجب عند تصميم الأسئلة أن توضع بترتيب تسلسلي منطقي، أو لا تسلسلي مقصود لدواعي التمويه المتعمد من أجل استجلاب إجابات صحيحة، ويفضل استخدام أنواع مختلفة من الخطوط بحيث تساعد على سرعة قراءتها واستيعاب ما فيها بأقل مجهود.

وكذلك يجب أن يؤخذ في الاعتبار متوسط الفترة الزمنية الكافية لإجراء كل مقابلة مع كل مفردة من المصادر محل الدراسة الاستخبارية، مع ضرورة إجراء مطابقة لما تم تخطيطه وتوقعه في البداية مع المتحقق بالفعل بحيث يتم تثبيت ذلك المتوسط أو تعديله وتصحيحه في الاستمارة النهائية.

وكذلك يجب ملاحظة الأسئلة الأكثر قبولاً وتحفيزًا للاستجابة لدى المصادر بحيث يتم البدء بها بتلك الأسئلة في الاستمارات النهائية وذلك باعتبارها مجموعات محفزة ومشجعة للمصادر على الاستجابة، أما الأسئلة غير المرغوبة فيمكن تأجيلها إلى نهاية قائمة التساؤلات أو نشرها بشكل بيني بحيث تحاط بأسئلة أخرى جاذبة يتوقع منها امتصاص أية مشاعر للتنافر وللعزوف من قبل المصادر واستعادة قبوله للاستمرارية في إجابة بقية الأسئلة.

وفيما يتعلق **بالمرحلة الثانية** المرتبطة بتحديد **أسلوب تصميم واختيار العينة** Method of Sample Design & Selecting or Sampling، فمن أهم الخطوات الواجب إتباعها في ذلك الصدد:

- التحديد الدقيق للفئة التي تقوم الدراسة على بحثها من بين الفئات المتعددة التي يحتمل تواجدها في نفس الدراسة. فمثلاً إذا كانت الدراسة تبحث أثر الأزمات الاقتصادية في بلد ما على انتشار الجرائم في مصر، تكون الفئة المنتقاة هي فئة المجرمين أو مرتكبي الجرائم، وهي ذاتها لابد من تحديد نوع منها بمزيد من الدقة. مثلاً تحديد فئة المرتشين في أجهزة الدولة، وهي التي لابد أيضًا أن تتحدد بمزيد من الدقة مثل التركيز على طبقة المرتشين في الأجهزة الحكومية الخدمية.. وهكذا

يظل الباحث في حالة تضييق لإطار الفئة أو الفئات التي تمثل بدقة أكبر وبتحديد أكثر دقة مصدر البيانات الأكثر ارتباطًا بالبحث.

- ثم يتم تحديد حجم الحالات المأمول استصدار البيانات من خلالها أو بكلمات أكثر دقة، تحديد حجم العينية Sample size: وكقاعدة عامة يجب على الباحث تحديد حد أقصى للفترة التي يمكن أن يقوم خلالها بعمل الدراسة الميدانية في حدود موارده المالية، وطاقاته وإمكانياته البشرية والمسموح الرسمي له به من قبل الجهة المشرفة أو الجهة المتحكمة في منح التصريحات اللازمة، كما يجب أن تؤخذ في الاعتبار عوامل أخرى ذات تأثير بشكل أو بآخر على ذلك التحديد مثل مدى صعوبة أو سهولة التعامل مع مصادر البيانات، ومدى صعوبة أو سهولة الإجراءات الرسمية المصاحبة أو مدى صعوبة أو سهولة الانتقال إلى ومن أماكن استجلاب البيانات ... الخ.

وهناك معادلة بسيطة يمكن أن يقرر حجم العينة الإجمالي بناء عليها وهي:

$$\text{ح}_{\text{ع}} = \text{م}_{\text{ص}} \times \text{ك}_{\text{م}}$$

أي حجم العينة المستهدف دراستها = متوسط عدد الحالات الممكن دراستها في اليوم الواحد × إجمالي عدد الأيام التي ستجري فيها الدراسة

وعند تحديد متوسط عدد الحالات اليومية محل الدراسة، يجب من خلال الدراسة الإرشادية حساب الفترة الزمنية المستغرقة في كل حالة والتعرف على السبب في تفاوت الفترات من حالة إلى أخرى مع تحديد تقريبي لصافي متوسط الفترة المستغرقة للتعامل مع كل حالة.

كما يجب مراعاة الأمور الأخرى التي يمكن أن تتداخل مع وقت الدراسة المذكورة مثل المسئوليات الأسرية والمهنية الأخرى، بل والمستلزمات الأخرى المرتبطة بالبحث ذاته كاستكمال تحصيل البيانات من مصادر أخرى على نحو ما أوضحناه سابقًا.

وكمثال مبسط، فإذا فرض أن إجمالي الوقت المتاح للباحث لإجراء الدراسة

الميدانية هي ثلاثة أشهر- يمكن تخصيص شهر منها للدراسة الاستطلاعية وللتعديلات المناسبة ولاستشارة المتخصصين والاستعانة بالمرجعيات المناسبة ولاستكمال الإجراءات الرسمية للحصول على التصريحات اللازمة وغيرها. فإذا ما اتضح من الدراسة الاستطلاعية أن متوسط المناسب للحالات اليومية محل التساؤل هو خمس حالات إذن بالتعويض في المعادلة سالفة الذكر:

يكون حجم العينة = 5 x 60 = 300 حالة

حيث (5) تمثل متوسط عدد الحالات اليومية و (60) تمثل إجمالي الفترة المتاحة للدراسة الميدانية بشكلها النهائي المنقح.

أما عن **تصميم العينة** فلابد من مراعاة الآتي:

- لابد أن تكون العينة المختارة غير متحيزة Unbiased، وممثلة لمجتمع الدراسة أفضل تمثيل، ومن أفضل الأساليب التي يمكن إتباعها لتحقيق ذلك الاختيار غير المتحيز هو «أسلوب تعدد المراحل» Multi-Stages Technique. فذلك الأسلوب يرتكز على قاعدة الانتقال التدريجي من الدائرة المركزية التي تحتوي على العينة التي سيتم دراستها بدون التمييز لمنطقة على أخرى.

ويتطلب اتباع أسلوب تعدد المراحل - بدوره - اتباع خطوات ثلاث أساسية هي:

1- تحديد المناطق التي ستتم دراستها من بين المناطق المختلفة للمجتمع المستهدف دراسته.

2- توزيع حجم العينة (الذي تم تحديده في الخطوة الرئيسية الأولى) «بين» between و«من خلال» within المناطق التي ستتم دراستها (وهي تلك التي تم اختيارها في الخطوة الفرعية الأولى[1].

3- سحب وتحديد الوحدات محل الدراسة The sampled cases.

---

(1) Atkinson, J. 1979, Reh E., 1981, pp 35-37.

ويتم ذلك بالاختيار العشوائي Randomly بين وحدات الدراسة المستهدفة وفقًا للعدد الذي يستوعبه حجم العينة الذي تم تحديده من قبل، وبحيث يتم توزيع العدد الإجمالي على عدد المناطق المختارة لتحديد عدد الحالات التي سيتم سحبها من كل منطقة مختارة بشكل منفصل وعشوائي كما ذكرنا.

والواقع أن عشوائية الاختيار يمكن أن تحدث بأكثر من طريقة أبسطها وأكثرها شيوعًا هو كتابة الحالات الإجمالية في قصاصات ورقية ثم سحب العدد المطلوب عشوائيًا من بينها. ولكن هناك طريقة أخرى تقوم على العشوائية المنظمة، كأن يتم اختيار عشوائي أولاً للمناطق محل الدراسة وذلك باستخدام اتجاه عقارب الساعة، وباستخدام ذات الاتجاه، يتم اختيار عشوائي للمنازل داخل كل منطقة مختارة، ثم تكرر الخطوة باختيار الحالات البشرية عشوائية داخل كل منزل مختار وهكذا.

والأكثر تنظيمًا أن يتم تقرير مؤشرات محددة ترتبط بعشوائية الاختيار، مثل الالتزام الدائم بالسير في اتجاه عقارب الساعة، ثم باختيار (ثاني وحدة تقابل الباحث في كل خطوة، فيتم اختيار ثاني منطقة مدونة في القوائم المرفقة ثم يتم ثاني حي موجود داخل المنطقة المختارة، ثم يتم اختيار ثاني شارع داخل الحي المختار، ثم يتم اختيار ثاني عطفة داخل الشارع، ثم يتم اختيار ثاني منزل داخل العطفة، ثم يتم اختيار ثاني شقة يصادفها الباحث بعد دخوله المنزل المختار. وداخل الشقة المختارة يمكن عمل اختيار عشوائي من خلال قصاصات الورق إن وجدت أكثر من حالة قابلة للدراسة وذلك طبعًا على سبيل المثال وليس التحديد أو الحصر.

وعلى ضوء العينة المختارة، وفي إطار أهداف الدراسة، يتم تنفيذ المرحلة الثالثة وهي إعداد استمارة الاستقصاء. ومن القواعد الهامة التي يجب أن تراعى في تصميم **استمارة الاستقصاء** المستخدمة في إجراء الدراسات الميدانية أن تبدأ بمقدمة أو ديباجة صغيرة a Preamble يقوم الباحث فيها بتعريف نفسه وبعرض فكرة مبسطة وموجزة عن الهدف من المقابلة، والتأكيد على سرية المتحصل عليه من إجابات، والتوضيح الدقيق بأن كل ما يتم تحصيله من بيانات إنما يؤخذ للاستخدام في

أغراض علمية تحليلية بحتة والإقرار بأنها لن تستخدم أية استخدامات أخرى، وبقدر الإمكان يفضل تجنب تسجيل أرقامًا مسلسلة أو رموزًا تصنيفية تتصف بصفة العموم[1].

ومن أجل الحصول على أفضل باقة من البيانات المستهدفة من خلال إجابات الوحدات محل الدراسة، يجب أن يراعى الآتي:

أ - استخدام الأسئلة «المغلقة» Closed في غالبية بنود الاستمارة وذلك بأن يوضع السؤال في شكل بدائل مختلفة يتم اختيار إحداها أو بعضها أو كلها، مع تجنب الغموض والتعقد الذي يسبب سوء الفهم للسؤال وبالتالي إعطاء إجابات غير صحيحة أو مضللة.

ب- في بعض الحالات، يفضل تقديم أسئلة «مفتوحة» Open بمعنى كتابة سؤال وترك حرية الإجابة الكاملة بألفاظ وعبارات كلها تقولها الوحدة محل التساؤل بحرية تامة. ويتم ذلك في حالات مناسبة مثل الرغبة في التعرف على الرأي العام في أمور عامة مصيرية كالدستور أو السعي إلى الاستفادة من خبرات فكرية أو عملية خاصة بالوحدات محل الدراسة.

ج- أما البيانات التي تتناول البنود الشخصية جدًا مثل الدخل الشخصي، فيجب أن يتم بشكل غير مباشر وبطريقة المدى الفئوي range كأن يسأل عن الدخل أين يقع في الفئات الآتية (أقل من 500 جنيه)، (500، 1000) جنيه، (أكثر من 1000 جنيه ..). كما اقترح «بارتين»[2] أن يتم توجه ثلاثة أسئلة للحالة الواحدة حول ذات البند أحدهما يدور حول المتوسط الشهري للدخل، والثاني يستفسر عن الدخل الشهري لأي فرد آخر في الأسرة يساعد على

---

(1) Wikan A., 1980.

(2) Parten M., 1965, p. 173.

الإنفاق على الأسرة، والثالث يتضمن أى دخل إضافي يرد إلى الأسرة بشكل شهري منتظم.

د- وكما نوهنا من قبل، يفضل توجيه مجموعة من الأسلة التكميلية التي يمكن أن تمثل مؤشرات غير مباشرة على مستوى دخل الحالات المعنية أو على أي بند آخر لا يرغبه مصدر البيان المباشر، وذلك مثل السؤال عن (المنفق الشهري على الدروس الخصوصية أو على الطعام والشراب أو على الكتب والصحف والمجلات) و (متوسط المدفوع كإيجار شهري أو كصيانة شهرية) و(متوسط المدفوع في فاتورة الكهرباء والغاز والمياه كل على حده)؛ وذلك لدواعي التحقق من صحة ما يقم التصريح به ومن مطابقة كل ما يقال.

هـ- ونظرًا لشدة الحساسية التي تتصف بها مثل تلك الأسئلة، تتبع بأسئلة أخرى محببة للحالات المعنية أو يتم تحليل تلك التساؤلات بعضها البعض الآخر، مثل سؤال الحالة عن أهم الهوايات أو عن الطموحات المأمولة أو عن الشخص المفضل والقدوة، ثم يتبع بسؤال حول المنفق الشهري على الأسرة ثم يتبع بسؤال خفيف حول رأيه في الأسعار السائدة ... وهكذا.

والواقع أن هناك الكثير الذي يجب ويمكن أن يضاف إلى ما سبق من إرشادات وضوابط، لكن تجنبًا للإصابة بالتشتت أو الملل نفضل الانتقال إلى الإشارة إلى المرحلة الرابعة من مراحل إعداد دراسة ميدانية التي تم ذكرها في جزء سابق، وهي تلك المرتبطة بعرض مجموعة التحفظات الواجب التنويه إليها بشأن ما قد يكشف البيانات المتحصلة من ثغرات أو قصور يمكن أن يتسبب في إحداث خلل في التحليل المزمع إجرائه باستخدامها، وذلك يفيد الباحث نفسه في تعويض تلك الثغرات بوسائل أخرى إضافية أو يمكن أن يفيد الباحثين الآخرين أو المستخدمين لتلك البيانات في اتخاذ اللازم من أجل المعالجة الصحيحة فيما يستجد من دراسات أو في استكمال الصورة ببيانات أخرى من مصادر أخرى.

ومن الأشكال الشائعة لما قد يحدث من ثغرات ونواحي قصور تستدعي التحفظ بشأنها نذكر الآتي:

- وجود عدة قياسات مختلفة باستخدام معايير مختلفة لذات النوع من البيانات مما يتسبب في تباين المتاح منها في المصادر المتاحة لإصدارها - وحتى البيانات التي يتم استجلابها من مصادرها المباشرة مثل تلك المتحصل عليها من خلال الدراسات الميدانية أو المقابلات الشخصية. فيمكن أن يختلط بيان الصادر منها مع طبيعية كل حالة، ووضعها المزاجي، ووعيها الإدراكي بأهمية الدراسة أو بأهمية ما يتم التصريح به، والحالة التعليمية أو الثقافية بل والاقتصادية لمصادر البيانات التي قد تكون في عجلة من أمرها عند تحصيل البيانات منها باعتبار وقت المقابلة محسوبًا بالسالب على الوقت المتاح لإنجاز أنشطة معيشية خاصة.
- عدم الدقة في توضيح أو فهم المصطلحات أو المؤشرات التي يتم بموجبها تجميع بيانات في أطر المعاني المختلطة أو المبهمة المشاعة حولها.
- عدم الكفاية الكمية للحجم اللازم من البيانات المطلوب استخدامها لتوصيف أو لتحليل موضوع ما أو عدم وجود سلسلة انتظامية متتابعة كاملة من البيانات المستهدفة مما يضطر الباحث إلى استكمالها من مصادر أخرى مختلفة المضامين المرتبطة بذات البيانات المعنية أو مختلفة القياسات كما سبق التوضيح.
- عدم سير الباحث على الخطة المعدة أصلاً أو عدم استخدامه لذات المناهج أو الأدوات المحددة فيها على الوجه الأكمل لأسباب قد ترتبط بمقاطعات قدرية أو بصعوبات إجراءات بيروقراطية مما يضطر الباحث إلى الاختزال في بعض الأجزاء.
- عدم توافر الوسائل أو الأدوات التجريبية أو التطبيقية الفعلية اللازمة لتحليل البيانات مثل برامج الكمبيوتر المتاحة أو ما يجب أن يتاح في الدراسات المعملية أو الحقلية أو ما شابهها.

وبطبيعة الحال فكل بحث وكل دراسة وكل مؤلف فكري علمي له مشاكله

التي ترتبط به، والباحث الماهر هو الذي لا يتعمد إخفاء نواحي القصور ويلجأ إلى فبركة البيانات والنتائج فيضلل بذلك الآخرين، لكنه دائمًا ما يتسم بما يطلق عليه «الأمانة العلمية» ويكون أول من ينتقد عمله ويوضح ما يتضمنه من محذورات تمامًا كما عليه أن يبرز ما يتسم به من مميزات طالما أنه بذل أقصى ما في وسعه في اتباع القواعد العلمية الصحيحة، وبذلك ما يستطيعه في معالجة الثغرات والأخطاء المحتملة بأفضل الوسائل التعويضية الملائمة.

وقبل الانتقال إلى الباب التالي المرتبط بوسائل تحليل البيانات، يجب استكمال الصورة الحالية بعرض مختصر لكيفية حفظ البيانات والمعلومات التي يتم تجميعها وكيفية إعدادها لتصبح جاهزة للاستخدامات التحليلية أو للعرض بالشكل المستهدف لها.

كيفية حفظ وإعداد البيانات والمعلومات وإعدادها للاستخدام النهائي

الواقع أن الباحث يلتزم بمهمة تمهيدية مزدوجة فيما يتعلق بتوفير المواد اللازمة للبناء الفعلي لبحثه وتلك المواد الخام تتمثل بشكل أساسي في مجموعتين أساسيتين، إحداهما البيانات Data المنقولة عن مصادر مختلفة أو المتحصلة بواسطة الباحث ذاته، والمعلومات Information المنقولة أو المشتقة من مصادر علمية أو فكرية مختلفة أو التي تم تجميعها أو استنباطها من قبل الباحث نفسه من خلال المشاهدات والمصادر المعرفية المباشرة كالظواهر الطبيعية والتجارب الشخصية وغيرها.

والواقع أن الاحتفاظ بالبيانات وبالمعلومات وإعدادها في فترة ما قبل الدراسة العملية أو التحليلية أو التطبيقية المستهدفة يختلف باختلاف هوية وطبيعة أو نمط المواد المعرفية المعنية. ويمكن أن يستعان بالشكل التالي للتعرف على صورة إجمالية لتوضيح كيفية تحقيق الهدف الجاري.

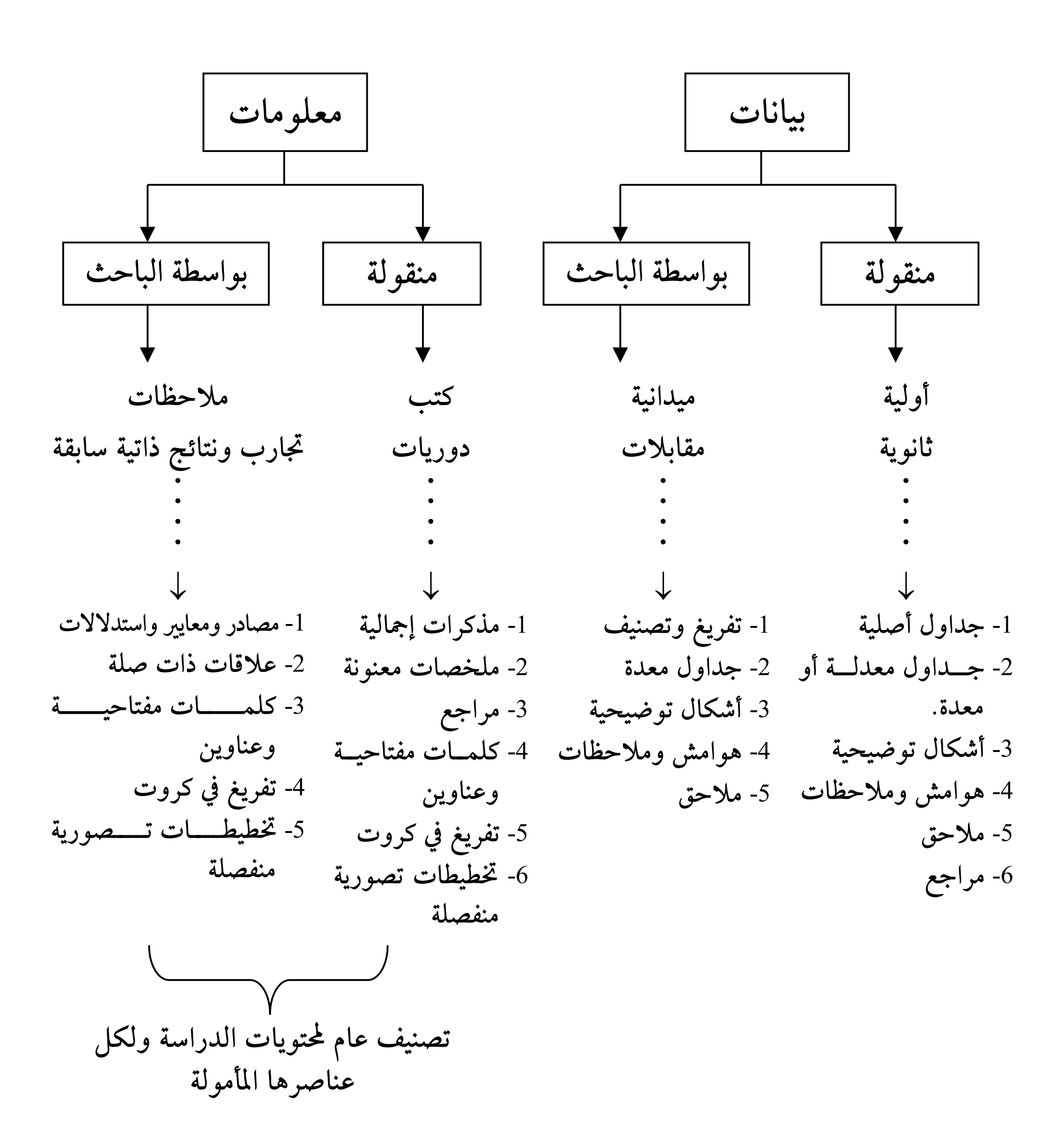
أولا: فيما يتعلق بالبيانات

كما يتضح من شكل (3) (الموضح في الصفحة التالية)، فالبيانات المزمع استخدامها في البحث يأتي بعضها بالنقل من مطبوعات ومنشورات لبيانات أولية أو ثانوية، وبعضها الآخر يمكن أن يتم تحصيلها بشكل مباشر من قبل الباحث ذاته من خلال الدراسات الميدانية والمقابلات الشخصية والمراسلات البريدية والاتصالات وغيرها.

(1) وفيما يتعلق **بالبيانات المنقولة** عادة ما يحتفظ بها الباحث في أحد أشكال الوحدات أو الاستخدامات الآتية:

- جداول يتم نقلها أو نسخها كما وردت في مصادرها الأصلية مثلما يحدث عند نقل جدول عن بعض الإحصاءات السكانية.
- إعداد جداول معدلة أو بشكل مختلف بما يخدم أهداف البحث بحيث تملأ تلك الجداول المعدلة أو الجديدة بمتناثرات مختلفة ما تم تجميعه من بيانات مع إضافة خانات يضيفها الباحث لاستكمال صورة ما قبل إضافة مختلطة ... الخ، وتلك الخطوة أفضل من سابقتها لضرورة لمس الشخصية الفكرية التحليلية الإبداعية المختلفة للباحث.
- وعلى نفس النهج يتم إدراج أشكال وخرائط بيانية وتوضيحية منقولة بحالها أو معدة بالكامل من قبل الباحث باستخدام المتحصل عليه من البيانات الملائمة أو منقولة مع إجراء بعض التعديلات والإضافة أو الحذف ..
- إدراج بيانات قد لا تصلح للاستخدام داخل متن البحث لكنها تفيد كهوامش أو كملاحظات يتوقع أن تفيد خاصة في حالة تفسير العلاقات السببية المختلفة وفي الربط بين نتائج الدراسة الجارية من ناحية ونتائج الدراسات المناظرة من ناحية أخرى.

شكل (3) إعداد البيانات والمعلومات للعرض وللتحليل التطبيقي



- إدراج بيانات ليس بغرض استخدامها في متن البحث لكن بغرض عرضها في جزء الملاحق الذي قد يكون خارج نطاق البحث المحدد سلفًا لكنه يفيد في استكمال الصورة الإجمالية التي تزيد من درجة التآلف مع البحث وعناصره، وتساعد على تحفيز آخرين للاستمرار على ذات الدرب - كلٌ بمنهج مختلف وبمعالجة استكمالية.
- تخصيص مدونة خاصة لكتابة تفاصيل المراجع والمصادر المنقول عنها كل ما سبق مع تحديد دقيق للصفحات الأصلية المستعان بها والالتزام بالطريقة العلمية المنضبطة.

(2) وفيما يتعلق بالبيانات **المستقطبة مباشرة بواسطة الباحث** من خلال الجولات الميدانية والمقابلات الشخصية والمراسلات البريدية والاتصالات الهاتفية أو الإلكترونية، فالأمر يختلف قليلاً في بداية الأمر نتيجة للوسائل المستخدمة في استقطاب البيانات بشكل غير مباشر مثل الأسئلة اللفظية التي ترتبط بإجابات لفظية لكنها توحي ببيانات يمكن تحويلها إلى رقمية كتلك المرتبطة بالأجهزة الكهربائية المتاحة والتي ينتهي تقييم قيمتها النقدية التقريبية إلى استنباط متوسط الدخل المتاح. وفي تلك الحالة يتم الاحتفاظ بالبيانات المتحصلة بالطرق الموضحة على النحو التالي:

- تفريغ البيانات من الاستمارات المخصصة لملء بنودها بالإجابات وبالبنود المستجلبة من مصادر البيانات.
- إعادة تجميع البيانات وتصنيفها كلٌ بما يتلاءم معه ثم وضعها في جداول معنونة بالمجال المرتبطة به ومرقمة بأرقام مسلسلة مكملة لسلسلة الجداول المعدة في الخطوة الأولى، أو توضع لها أرقامًا مسلسلة مستقلة لحين إعادة الترقيم الإجمالي وفقًا للشكل الكلي لعرض البحث إجمالاً.
- إعداد أشكال بيانية وتخطيطية ومسودات توضيحية لما تشير إليه البيانات المتحصل عليها.
- إعداد جزء منفصل للهوامش والملاحظات المرتبطة بالدراسة الاستقصائية العملية لاستخدامها عند الحاجة.

• إعداد جزء منفصل تدرج فيه كل متبقيات حصيلات الدراسة الاستقصائية مع إعطائها تصنيفات بديلة تمهيدًا لإدراجها داخل متن البحث لكن في صورة ملاحق تستكمل بها صورته التوضيحية.

ثانيا: فيما يتعلق بالمعلومات

فهناك أيضًا قسمان تستجلب المعلومات وتستنبط بناء عليهما؛ وكل منهما يمكن أن يتم ادخاره في الأوعية المناسبة وبالأسلوب الملائم الذي تسهل معه استرجاع البيانات ووضع كل منهما في وضعها الصحيح داخل من البحث في العرض النهائي المقدم للأوساط العلمية والمهتمة.

(1) فيما يتعلق بالمعلومات **(المنقولة)** من مطبوعات ومنشورات علمية تتوافر فيها الشروط المذكورة في جزء سابق:

• يقرأ كل مطبوع من الكتب أو الدوريات أو غيرها أولاً على ثلاث مراحل .. **قراءة إجمالية تصفحية** بدءً بقراءة العنوان ثم المقدمة التعريفية بهن ثم الخاتمة والاقتراحات التي تلخص حالته النهائية، ثم قائمة المحتويات التي تقدم صورة متكاملة لمكوناته وتعطي للقارئ دليلاً لعناوين كثيرة يمكن استنباط عنوان مناسب منها للبحث المزمع إنجازه، أو الاستعانة بها للاقتداء في البند المقابل بالبحث المستهدف. ثم يترك الباحث المطبوع جانبًا ويقوم بتدوين ما تذكره مما قرأ، وبكتابة انطباعاته الفكرية والاستقرائية حول ما فهم منه.. وبذلك تصبح لديه **مدونة عامة** تخصص لتلك المرحلة الاستقرائية الأولى.

• في مرحلة تالية يستعين الباحث بما يمكن أن يطلق عليها **مدونة للملخصات الهامة** تكون تلك المدونة عبارة عن كشكول كبير مثل كراسات الرسم التي تتميز بكبر مقاييس أطوال أوراقها ومساحتها الإجمالية بحيث تكون كل ورقة صالحة لاستيعاب تصور كامل يدونه الباحث حول كل جزئية منفصلة من المعلومات التي ينقلها من المطبوع مع ربطها بالأبعاد والجوانب المختلفة التي ترتبط بها

ورسم **شجيرات معرفية** تحمل كل شجرة عنوانًا يدل على مكوناتها، وفي الجذع والفروع والتفريعات يتم تدوين كل معلومات ذات أهمية خاصة في الوضع الملائم لها، وذلك من خلال قراءة ثانية لذات المطبوع مع تدوين كل النقاط الهامة الموجودة في المطبوع والتي يعتقد أنها ستفيد الباحث بشكل أو بآخر فيما بعد بحيث تصبح تلك المدونة هي البديل المتكامل الذي يغني الباحث عن الرجوع ثانية إلى المطبوعات الأصلية، وتصبح كبسولة معرفية استقرائية مركزة يتيسر اللجوء إليها عند الحاجة طوال فترة إعداد البحث بمراحله المختلفة.

- قبل التخلي عن المطبوعات المنقول عنها المعلومات المذكورة، يفضل كتابة قائمة المراجع والمطبوعات التي تمت الاستعانة بها في مدونة مخصصة فقط لكتابة **المراجع** بحيث تكتب بكل التفاصيل المرتبطة بها والتي ستنقل بحالها للتدوين في نهاية البحث بعد الانتهاء التام من إعداده. لكن في تلك المرحلة تتم إضافة بيانات كاملة حول أماكن تواجد تلك المطبوعات وأرقام تصنيفها (في حالة أن تكون قد تمت استعارتها من مكتبات معنية مثلاً) وذلك لتيسير ولضمان العثور عليها إذا كانت هناك حاجة مستقبلية للإطلاع عليها مرة أخرى.

والواقع أن تلك المرحلة يهملها كثير من الباحثين ويتناسونها مفضلين أن يقومون بتدوين المراجع في وقت لاحق، لكن انشغالهم بمراحل البحث المختلفة يمكن أن ينسيهم كثيرًا مما استخدموه من مراجع أو ينسيهم أماكن تواجدها مما يتسبب في حالة ارتباك كبيرة تضيع الكثير من الأعصاب والوقت والمجهود، بالإضافة إلى التعرض لانتقادات من قبل المحكمين وإلى القراء المتخصصين وصفهم بعدم الالتزام بمبدأ «الأمانة العلمية» الذي يعتبر ميثاق شرف لكل باحث والذي يقضي بحتمية توثيق كل ما يكتب من معلومات وبيانات منقولة وإرجاع فضل تقديمها إلى المصادر الأصلية التي قامت بذلك في الواقع.

لذا يفضل إعداد مدونة خاصة بالمراجع من بداية مرحلة الاستعانة بها وتدوينها بتفاصيلها أولاً بأول.

- يمكن أيضًا تخصيص مدونة للمصطلحات بترجماتها وبمعانيها المتوافرة في المطبوعات المستعان بها بحيث يقوم الباحث بإعداد **قاموس تخصصي شخصي** يستخدمه عند الحاجة، ويخصص الباحث فيه جزءًا لتدوين عبارة جميلة أو متميزة تمت قراءتها في المطبوعات المختلفة وذلك بهدف إثراء حصيلته اللغوية وزيادة مقدرته وكفاءته في مرحلة الكتابة النهائية ببحثه. وفي ذات المدونة يتم أيضًا تخصيص قسم للمترادفات اللغوية العربية والأجنبية (إنجليزية، فرنسية، ألمانية ... الخ) ويكتب بجانبها عدد حروف كل كلمة حيث يفيد ذلك في حالة الاضطرار إلى تخفيض عدد الصفحات المكتوبة، فيكون من إحدى طرق ذلك هو استخدام كلمات أقل عددًا في حروفها من مترادفات أخرى لها. وتجدر الإشارة هنا إلى أن كثير من الدوريات خاصة الأجنبية منها (مثل المطبوعات الدورية لمنظمة الصحة العالمية) تشترط لقبول البحث أن لا يتعدى عددًا إجماليًا من (الحروف) Letters أو من الكلمات Words وليس من الصفحات كما تشترط دوريات أخرى مناظرة خاصة في الدول النامية. فإذا ما درب الباحث نفسه على أخذ عدد حروف كل كلمة في الاعتبار فسوف يكون أكثر قدرة على التحكم في ما يكتبه وعلى عرض المعلومات بأقل قدر من الكلمات التي تتيح فرص أفضل للتركيز ولاستيعاب المعروض والتي تحقق المعادلة الصعبة (إنجاز + إيجاز أو إنجاز بإيجاز) done briefly. ويمكن في ذات القاموس اللغوي تخصيص جزء آخر للمتقابلات اللغوية حيث يستفاد من ذلك في عرض متنوع الألفاظ العمومية بما يقلل من درجة الملل من قراءة ذات الألفاظ وتكرارها المستمر.. اللهم إلا إذا كانت تلك الألفاظ تتعلق بمصطلحات علمية محددة ففي تلك الحالة لابد من الالتزام بتوحيد استخدامها كما هي بدون أي تغيير أو تبديل أو ترديف.

- إعداد قصاصات ورقية مربعة أو مستطيلة أو الكروت المعتاد استخدامها لأغراض حفظ المعلومات والتي تباع في المكتبات المختصة بالأدوات المدرسية والتعليمية، حيث يتم تدوين **كلمات مفاتيح key words** تشير باختصار ورمزي

إلى أهم المواضيع الأساسية والفرعية والمتفرعة المزمع التعرض لها في البحث بحيث تدون هي الأخرى في شكل خرائط تصنيفية توضيحية يمكن أن يستدل بها الباحث طوال فترة دراسية.

- إعداد ما يمكن أن يسمى بـ **مجموع الحقائق والملاحظات** وذلك بتفريغ ما قد يعتقد الباحث بأنه يخرج عن نطاق الدراسة لكنه قد يفيد في حالات كثير مثل شرح وجهة نظر حول نقطة بحثية ما والاستدلال المنطقي بأمثلة أو بدراسات مختلفة عن النطاق محل الاهتمام، وكذلك يمكن نسخ فقرات بعينها (بين علامتي تنصيص) لما قد يكون لها من أهمية. وبدلاً من استخدام الكشاكيل كبيرة الأطوال والأحجام يفضل هنا استخدام بطاقات ذات حجم موحد (4×6 بوصة أو 5×8 بوصة مثلاً حسب الرغبة) بحيث يمكن حفظها وسحب المرغوب منها كلٌ في حينه. ولابد من كتابة بيانات المرجع بالكامل أو إعطائه رقمًا مسلسلا في المدونة الخاصة بالمراجع، وكتابة ذلك الرقم المسلسل في كل كارت يحوي معلومة تم اقتباسها منه.
- إعداد **مجمع للمخططات وللخرائط وللأشكال التوضيحية** بشكل مبدئي بحيث يمكن الاستعانة به كل في حينه وبحيث يمكن تعديل كل مخطط أو شكل أو جدول وقت تدوينه أو تضمينه داخل متن البحث.

(2) وفيما يتعلق بالمعلومات التي يستجلبها الباحث أو يكونها بنفسه وبأدواته الخاصة من خلال المشاهدات والتجارب الذاتية وغيرها، فيمكن الاحتفاظ بها في أقسام مختلفة كالتالي:

- مدونة تسجل فيها المصادر غير التقليدية أو الذاتية التي تم الاعتماد عليها في تكوين المعلومة (مثل استخدام حاستي الشم والبصر لاستنتاج استمرار ظاهرة السحابة السوداء الظاهرة في منطقة البحث أمام الباحث).
- مدونة تسجل فيها علاقات مختلفة بين ظواهر معينة أو أنشطة ما أو سلوكيات ما

لاحظها الباحث بتتبع دائم لها بحيث يمكنه بعد ذلك استخدامها كفرضيات بحثية للتحقق من صحتها أو من عرضها أو من تأقيت حدوثها.

- كروت مختلفة اللون عن المستخدمة في تدوين البيانات أو في تدوين المعلومات المنقولة من مراجع، حيث يتم استخدام لكتابة كلمات وعناوين مفاتيح تصورية بمجهود الباحث الشخصي ومستنبطه من مشاهدات ومما يمكن أن يلهمه الله به من أفكار وإبداعات تطرأ على ذهنه أثناء فترة الإشغال الذهني الدائم بموضوع البحث.
- وعلى نحو ما سبق، يقوم الباحث بتحديد تخطيطات تصورية منفصلة في قصاصات أو كروت منفصلة لاستخدامها وقت الحاجة.
- وعلى ضوء كل ما تقدم وبالاستعانة بذلك يمكن للباحث إعداد تصنيف عام لمحتويات الدراسة وعرض جدول زمني تقديري لإنهاء كل جزئية من تلك المحتويات في إطار الوقت الكلي المسموح به.

ومهما أفضنا في التفاصيل، تحكمنا دائما المحدودية النسبية التي تدعونا الآن إلى التوقف عند هذا الحد آملين أن يكون كافيًا لنفع الباحث بالاسترشاد بعناصره، لكن على الباحث إن اكتشف طرقًا أفضل أو إضافية أن يتبعها بتصرف - فما نقدمه يمثل خلاصة تجاربنا وتجارب كثير من العلماء المختصين، وعلى الباحث أن يدلي بدلوه في ذلك المضمار في كتاب يخص بعد أن يصبح هو الآخر خبير محنك في ذلك المجال البحثي الإرشادي.

وحتى لا يصبنا التيه، نود أن نتذكر معًا ما سبق عرضه في البابين الذين انتهينا من عرضهما توًا. فقبل إرشاد الباحث إلى خطوات البحث العلمي وكيفية إعداده، كان من الضروري أن نرشده إلى فهم هويته التي يجب أن يكون عليها حتى يصبح مقبولاً كعضو داخل الدوائر العلمية الراقية، وكذلك أن نقدم له بطاقة تعارف مع المجال الذي يطمح إلى الانتساب إليه، وقد حدث ذلك في الباب الأول التعريفي

بالباحث العلمي، وبالبحث العلمي. وباعتبار أن الخطة البحثية هي المفتاح الأوحد الذي يجب على الباحث امتلاكه للولوج الفعلي إلى دنيا البحث العلمي المأمول، تم الانتقال إلى الباب الثاني لتوضيح كيفية إعداد الخطة البحثية بكل المقدمات التمهيدية اللازمة لإعدادها، ثم بكل العناصر المكونة لها، ثم بكيفية تنفيذ الإجراءات المواكبة معها من استجلاب للبيانات وللمعلومات وإدخارها بالشكل الملائم لاستخدامها عند الحاجة إليها.

ومن بين ما تعرفنا عليه خلال العرض السابق، أهم المناهج العلمية التي يمكن للباحث أن يتبع أحدها أو بعضها أو كلها.

ويتبقى الآن التعرف بمزيد من التعمق على الأدوات التحليلية التي يجب على الباحث استخدامها عندما يتبع مثل تلك المناهج من أجل الاستخدام المستهدف للبيانات وللمعلومات التي انتهينا توًا من الحديث عنها وحولها، لنذهب معًا إلى الباب التالي لنعرف ذلك.

الباب الثالث

بعض الوسائل والأدوات الوصفية والتصنيفية والتحليلية

يحتوي هذا الباب على الفصول التالية:

الباب الثالث

بعض الوسائل والأدوات الوصفية
والتصنيفية والتحليلية

مقدمة:

الغالبية العظمى من الباحثين خاصة الناشئين أو غير المتخصصين أو المتدربين منهم دائمًا ما يختلط عليهم التمييز اللازم بين مناهج البحث العلمي من ناحية، وبين أساليب وأدوات التحليل العلمي من ناحية أخرى. فالمنهج العلمي أو المنهجية العلمية Methodology or Methodism يعني طريقة معينة للتفكير وللتحليل العقلاني والمنطقي ويقوم على فلسفة مذهبية محددة المعالم والاتجاهات. وبكلمات أخرى هو طريق معين يختاره الباحث لتنفيذ بحث علمي بكل أهدافه ومكوناته الأساسية الأخرى. أما الأسلوب العلمي Method فهو يمثل الأداة Tool التي سيقوم الباحث باستخدامها داخل الطريق العلمي الذي اختار أن يسير على خطاه لكي ينجز بحثه العلمي على النحو الملائم.

وليس بشرط أن تكون كل الأدوات أو الأساليب العلمية المستخدمة متطابقة

كمًا أو كيفًا بشكل موحد ومناظر مع المناهج العلمية التي اختارها الباحث. فعلى سبيل المثال، قد يختار الباحث المنهج التاريخي والمنهج الوصفي لدراسة وضع اقتصادي ما في بلد ما، كأن يقرر دراسة تطور التصنيف الائتماني لمصر منذ فترة التسعينيات (حيث تم تطبيق سياسة الإصلاح الاقتصادي) حتى ما بعد ثورة 25 يناير 2011. في تلك الحالة. ترتبط الدراسة بالمنهج التاريخي من حيث أخذ البعد الزمني في الاعتبار، وتبرير النطاق الزمني الذي تم تحديده هنا نتيجة لتتابع علاقات تأثيرية تسببت في إحداث تقييمات مختلفة للتصنيف الائتماني موضع التركيز البحثي. غير أن علم الاقتصاد لا يكتفي أبدًا بمجرد تقديم عرض وصفي لما هو كائن بالفعل من تطورات وتأثيرات معينة، لكن لابد أن تستكمل الصورة بإجراء دراسة تحليلية لكل ما حدث من أجل تقديم إضافة جديدة تبرر نشوء البحث والقيام به من منظور اقتصادي - تلك الإضافة هي التي يمكن التعرف عليها من خلال قراءة خطة البحث وخصوصًا من خلال الهدف المقرر، والأهمية الموعود بحدوثها من إنجازه. وعادة ما ترتبط أساليب البحث وأدواته التحليلية بهذين البندين على وجه الخصوص. فعلى سبيل المثال، إن كان الهدف من البحث هو التعرف على الآثار الاقتصادية المختلفة على التصنيف الائتماني لمصر، يمكن أن يتقرر أسلوب الارتباط والانحدار المتعدد كأداة تحليلية ملائمة للتوصل إلى تحديد والي قياس مثل تلك الآثار وذلك باستخدام بعض المعايير القياسية المناسبة التي يتم باستخدامها تحديد المتغير الممثل للتصنيف الائتماني (كمتغير تابع) من ناحية، وتلك التي بموجبها تتحدد مجموعة المتغيرات التفسيرية التي سيتم استخدامها باعتبارها العوامل الأكثر تأثيرًا على ذلك المتغير التابع، ثم تجري بعد ذلك الخطوات القياسية التحليلية اللازمة.. وكل ذلك سيحدث داخل إطار المنهج التاريخي الوصفي.

وقد يقول قائل لماذا لا يقول الباحث بأنه استخدم المنهج الإحصائي في حالة استخدام الوسيلة الإحصائية المذكورة؟ فنجيب عليه بأنه ذلك يحدث إن كان البحث كله قد قرر أن يدخل بكثافة وبتعمق داخل الطريق الإحصائي المكثف.

لكن في مثالنا التمهيدي هذا، يقوم الباحث فقط باستخدام الأداة الإحصائية بشكلها الضروري المبسط لأداء مهمة ضمنية داخل إطار بحثي متعمق في البحث التاريخي التطوري بصفته. إنه مثل من تعد كيكة أسفنجية أو تورتة.. فالمسميات مختلفة لاختلاف درجة الاهتمام والتركيز على الهيئة الإجمالية لكليهما، برغم أن التورتة هي في أصلها كيكة أسفنجية.

وبغض النظر عن إشكالية التمييز السابق الإشارة إليها بين المنهج العلمي والأسلوب التحليلي فلنذهب معًا الآن للتعرف على المزيد عن أساليب البحث العلمي والأدوات ذات الصلة والأكثر شيوعًا واستخدامًا.

الفصل الأول

المنظور النظري والكمي

مقدمة:

قبل البدء في استعراض الأساليب المعنية، يجدر التنويه أولاً إلى حقيقة هامة لابد من الانتباه إليها في هذا الصدد. فهناك اعتقاد سائد بأن كلمة «وصفي Descriptive» دائمًا ما يرتبط بما هو كائن برمته وعلى حاله. لكن ذلك المفهوم يمكن أن يتضمن معان وأبعاد تحليلية أيضًا إذا ما تم إضفاء تعديلات أو تغييرات على توصيف ما هو كائن، وإذا ما اقترن ذلك التوصيف بعملية تفنيد وقياس للكائن بشكل يفصله عن حاله الأصلي محل الوصف إلى وضع وصفي تصنيفي مختلف الهيئة والملامح نتيجة لإعمال المشرط الفكري المتعمق وتغلغله في صلب مكوناته. ذلك المشرط الفكري قد يقوم بعملية مشابهة لما يفعله الجراح الماهر، لكنه يختلف في هويته ومواصفاته حيث يأخذ هنا وضع المعايير القياسية Standards أو المؤشرات الاستدلالية القياسية Indicators أو ما شابه هذا وذاك.

فعلى سبيل المثال، إذا ما تناول باحث فئة ما بالدراسة وذكر من مواصفاتها أنهم مصابين بسوء التغذية، نقلاً لتلك الصفة من دراسات أخرى متخصصة في

مجال التغذية، أو من أية مصادر أخرى ذات صفة وارتباط، فيكون ذلك عرض وصفي بحت وتكن الأداة المستخدمة لذلك العرض المنهجي الوصفي هي «النقل» أو النسخ المقلَّد Copycatting. لكن قد يرفض الباحث تطبيق أسلوب المحاكاة البحتة لفكر الآخرين، ويسعى إلى تعلم واستخدام المعايير المتخصصة المختلفة لحالة سوء التغذية مثل معيار «واترلو» القائم على مقارنة الوزن بالطول، ومعيار «جوميز» القائم على مقارنة الوزن بالنسبة للعمر، ثم يتبع التصنيف التخصصي الذي يوضح أن سوء التغذية يتضمن ثلاثة أنواع كل منها يشير إلى درجة مختلفة من ذلك المرض (خفيف) لا يتطلب إلا تحسين المتعاطي لما يتناوله من مواد غذائية؛ و(متوسط) أكثر خطورة يستدعى علاج من قبل خبير تغذية متخصص، و(خطير) قد يتسبب في الوفاة مما يتطلب العلاج تحت رعاية مركز بإحدى المستشفيات التخصصية.. في تلك الحالة لن يعرض الباحث الحالة الغذائية المعنية وفقًا لوحدة الآخرين لها، لكنه سيعمل مشرطه المعياري المعرفي ليحدد بدقة أي صنف وأية ظروف يتحدث عنها من خلال تحليله للحالة، ومن ثم يقال أن الباحث قد استخدم أداة تحليلية وصفية، حيث قام أولاً بتحليلها باستخدام القياسات الملائمة، ثم عرض الحالة ووصفها وفقًا لما خرج به من نتائج.. وكل ذلك بالاستعانة أيضًا بالدراسات الأخرى المماثلة.

وبناءً عليه فعندما نتحدث عن الأدوات الوصفية أو التحليلية هنا فنحن نركز على ما يتم إعمال العقل والفكر والتفنيد والقياس معه وإلا انتفت صفة «العلمية» عن طبيعة البحث المنجز.

1/1/3 عرض تصنيفي للمجموعات التخصصية في مجالات الوصف والتصنيف والتحليل العلمي

عندما نتحدث عن أساليب وأدوات التحليل الوصفية والتقييمية ونبحر في دهاليزها ستكتشف أن تحديدها أمرًا ليس بالهين، فهو يتطلب بالدرجة الأولى إعدادًا

تأهيليًا خاصًا للباحث يبدأ منذ مراحل التعليم الانتظامية حيث كان يدرس مثل تلك الأساليب والأدوات من خلال مناهج ومقررات مختلفة المسميات برغم تشابك وتداخل الكثير من مضامينها، مرورًا بما تتم دراسته في مراحل إعداد الباحث خلال الدراسات التمهيدية العليا المنتظمة التي تستمر على مدى عام أو عامين قبل أن يبدأ في إعداد بحثه المرتقب. لكن بالإضافة إلى كل ذلك، يجب على الباحث المرور بمزيد من التعمق بين الكتب المتخصصة في عرض الوسائل التحليلية مع الدراسات التحليلية مع الدراسات الأخرى المتصلة بالموضوع أو بالمجال البحثي وذلك من أجل التآلف والاندماج الكافي في دنيا التحليل العلمي التطبيقي إلى الحد الذي يمكن الباحث من المفاضلة بين الوسائل المختلفة واختيار الأفضل والأكثر ملاءمة منها لموضوع بحثه بكل متضمناته.

وفيما يلي بعض الوسائل التحليلية التي يمكن أن يستعين الباحث بإحداها أو ببعضها من أجل استكمال الجزء التطبيقي أو العملي القائم عليها.

أولا: الوسائل الإحصائية Statistical Methods

وهي التي يتم الاستعانة بها من خلال علم الإحصاء الذي يقوم على المعالجة الرقمية والكمية للأرقام والخروج بدلائل قياسية ومعيارية وبعلاقات قياسية عالية الدقة بين العوامل محل الدراسة والتحليل. ويقوم علم الإحصاء على فلسفة معينة توضح أهمية الاستعانة بأدواته القياسية التحليلية وهي أن الأمور إذا ما تم قياسها والتعبير عنها بالأرقام تصبح أكثر قابلية للتعرف عليها بدقة أكبر وبوضوح أكثر مما إذا ما كانت صعبة القياس الرقمي وبالتالي قد يؤدي التعبير عن هويتها وحقائق وضعها أمر أكثر غموضًا وأكثر عرضة للتشكك في مصداقيته

"when you can measure what you are speaking about and express it in numbers, you know something about it, but when you can not express it in numbers, your knowledge is a meager and unsatisfactory kind"

وبكلمات أخرى مختصرة، فإن التحليل الكمي القائم على استخدام الأرقام يعطي نتائج أكثر دقة ومصداقية ووضوحا عن التحليل الكيفي quantitative[1].

ومن أهم **الوظائف أو التطبيقات** التي يقوم عليها علم الإحصاء:

1- الوصف الكمي Numerical descriptive وذلك من خلال العرض الرقمي أو التصنيفي أو الترتيبي للبيانات لتوضيحها بصورة أكثر دقة وأكثر قابلية للتوضيح وللمعالجة التحليلية الصحيحة وذلك مثل وصف حالة البطالة في مصر من خلال عرض متغيرات تمثلها وأهمها عرض عدد الملتحقين بالقوة العاملة ولكنهم لا يستطيعون الحصول على فرص عمل حقيقية، وعرض أطوال فئة معينة عن الشعب مقاسة بالسنتيمتر، وعرض الأوزان مقاسة بالكيلوجرامات وعرض الأعمال مقاسة بالأيام (لحديثي الولادة) أو بالشهور (في فترة الرضاعة) أو بالأعوام (للبالغين).

كما تستخدم القياسات الرقمية لأغراض أخرى غير مباشرة ولا تتعلق بعرض قيم فعلية للمتغيرات محل الدراسة، لكنها تستخدم لأغراض ترتيبية Ordinal (كعرض الأرقام التصاعدية أو التنازلية للدرجات التقييمية لطلاب فرقة معينة)، وكذلك لأغراض تصنيفية أو تمييزية للبنود غير القابلة للقياس الرقمي مثل المهنة، والنوع، والجنسية، والحالة الاجتماعية ...الخ).

2- التفسير Inferential: ويقصد به إعطاء معنى ملموس للبيانات، وتوضيح أسباب الأحداث والعلاقات بين المتغيرات المختلفة.

3- تحديد طبيعة العلاقات السببية ومدى التقارب أو التباين بين النتائج أو القياسات أو الشواهد المختلفة وما شابه ذلك باستخدام وسائل مختلفة مثل الارتباط البسيط Simple correlation، الارتباط والانحدار المتعدد Multiple Regression

---

(1) Norman K. Hendersan, p. 2

Analysis، قياسات التشتت أو الانحراف Dispersion، المتوسط Average، الوسيط Median، قياسات النزعة المركزية Measures of central tendency، المنوال Mode... الخ.

4- نظرية المعاينة أو استخلاص النتائج بأسلوب قياسي Theory of sampling.

5- اختبار الفروض Testing Hypothesis.

6- اختبارات التحقق من النتائج التحليلية (مثل التباين Variance، النسبة الحرجة Critical Percentage.

7- عرض الجداول والأشكال والخرائط البيانية.

8- الاحتمالات.

وللأهمية البالغة التي تحظي بها الوسائل الإحصائية في المجال الوصفي والتحليلي لكل العلوم الطبيعية والاجتماعية تقريبًا، فسوف نتناولها بمزيد من التوضيح المختصر عبر الصفحات التالية، ولكن سوف يتم ذلك العرض التوضيحي بعد استكمال عرض بقية الوسائل الشائع استخدامها في المجال الوصفي التحليلي العلمي.

ثانيًا: الوسائل الرياضية Mathematical Methods

وهي المشتقة من علم الرياضيات العام، ومن بعض فروعه المتخصصة مثل الاقتصاد الرياضي Economic Mathematic. وعلم «الرياضيات» يختلف عن علم الرياضة، فالأخير يقوم على دراسة الحركة البدنية للإنسان، وإن كان له نفع في مجالنا هذا، فسيكون في التدرب على التمرينات البدنية والتمرينات الذهنية التي تساعد على تنشيط الدورة الدموية للباحث وعلى تنشيط العقل وملكات الحس والتفكير مما يساعد بدوره على زيادة درجة الكفاءة الذهنية التفكرية والتحليلية والإبداعية للباحث. إلا أن علم الرياضيات الذي يعنينا هنا يقترب بهويته كثيرًا من علم الإحصاء لأنه مثلا يقوم على التعامل مع الأرقام، ويركز على التحليل والوصف الكمي بشكل أكثر عمقًا واتساعًا، ويمكن القول بأنه علم «الدراسة

المنطقية» لكل الأشياء ولكيفيتها ولعمليات الربط بينها من خلال معادلات أو نماذج أو ما شابه، كما أنه يمثل علم الدراسة المجردة البحتة التسلسلية كاملة الحيادية للقضايا وللأنظمة.

ومن أكثر **المجالات والاستخدامات والتطبيقات** لعلم الرياضيات

1- حل المعادلات الآنية التي تمثل في صيغ رياضية تقوم على المساواة، والأحرف فيها تمثل القيم المجهولة المطلوب إيجاد قيم لها من خلال تلك المعادلات ويدخل في ذلك مجال بناء وتطبيق النماذج Modelling.

2- إجراء العمليات الحسابية والجذور والتركيبات والتربيعات ... الخ.

3- تحليلات هندسية مثل دراسات خواص الأشياء والعلاقات بينها، والأشكال من مربعات ودوائر ومستطيلات وغيرها مما يدخل في نطاق «الهندسة المستوية»؛ ودراسة الأشكال ذات الأبعاد الثلاثية مثل الكرة والمكعب التي تتم دراستها في مجال الهندسة الفراغية.

4- حساب المثلثات وحساب التفاضل والتكامل والاحتمالات ... الخ.

5- الجبر الذي كان المسلمون لهم سبق الريادة فيه مثل «الخوارزمي» و«ابن بدر».

6- الحوسبة، وذلك بحل المسائل الرياضية وبناء النماذج وتكوين المعادلات بعلاقات مختلفة وإجراء غيرها من العمليات من خلال الحاسوب (الآلة الحاسبة Calculator) أو الحاسب الآلي (الكمبيوتر Computer) باستخدام برامج مختلفة وفقًا للأغراض الموجهة لاستخدامها، مثل استخدام برنامج +SPSS للدراسات التحليلية الاجتماعية خاصة المقطعية منها Cross table وبرنامج STP في حالة السلاسل الزمنية التي تعالج سلسلة من البيانات عبر عدد ممتد من الوحدات الزمنية (أشهر، سنوات، ... الخ).

ثالثا: الوسائل القياسية Metrology أو المترولوجيا

وهي بالغة الأهمية في عمليات القياس المختلفة للأشياء وللمتغيرات الممثلة لعناصر التحليل ولمكونات المعادلات المختلفة ومثل تلك العمليات القياسية تستلزم ربطها بتحرير نسبة الخطأ المرتبة على القياس المستخدم لدواعي الدقة والتحرز. ومن الأمثلة على تلك العمليات القياسية: قياس درجات الإشعاع (مثل الضوء) وقياس الكميات الكهرومغناطيسية، وقياس درجات الحرارة، وقياس درجات الأمن الصحي والبيئي، وتقييم أداء النظام الضريبي، وتحديد مستوى الإنتاجية في نشاط أو مجموعة أنشطة ما ... الخ.

أما في العلوم الشرعية، فإن القياس يعتبر ليس فقط مجرد فرعًا علميًا عاديًا، لكنه علم بالغ الأهمية يقوم على عمليات القياس باستخدام القوانين الإسلامية من أجل التعامل الشرعي الصحيح مع الأمور المعاصرة والمستجدة والتي لا توجد فيها نصوص شرعية صريحة بشأنها ويسمى علم القياس الشرعي (Analogical deduction of Islamic Laws) وجدير بالذكر أن القياس الشرعي يمثل عمليات اجتهاد Dillgence تتطلب توافر خصائص معينة في الباحث الشرعي بحيث لا يسمح لسواه بالقيام بتلك المهمة[1].

ومما يجدر التنويه إليه، أن علم القياس الشرعي الذي يقوم على الاجتهاد المنضبط بالشروط الشرعية الصحيحة يتطلب من المستخدم لأدواته ولتوجيهاته الدقيقة أن يكون ملمًا إلمامًا كاملاً وكافيًا بخمسة علوم:

1- **علم القرآن** (الناسخ والمنسوخ، والمجمل والمفسر والخاص والعام، والمحكم والمتشابه، والكراهة والتحريم، والإباحة والندب ..)؛

2- **علم السنة النبوية** (معرفة الصحيح والحسن والضعيف والمسند والمرسل، ومعرفة

---

(1) زينب صالح الأشوح، 1997، ص 185-187.

ترتيب السنة على الكتاب وبالعكس باعتبار أن السنة تمثل بيانًا توضيحيًا وتفسيريًا لما ورد بكتاب الله والتعرف على كيفية ما قد يبدو من اختلاف بين نصوص بعض الأحاديث النبوية من ناحية وبين آيات قرآنية مناظرة، والتمييز بين الأحاديث النبوية والأحاديث القدسية...)؛

3- **علم اللغة العربية** (العربية تحديدًا) (كمعان، ومجازات، وقواعد نحوية وغيرها ما يمكن الباحث من التفسير الصحيح لمعاني القرآن والسنة)؛

4- **العلوم الفقهية** وأقاويل علم السلف من إجماع واختلاف؛

5- **علم القياس** (الذي يفيد أساسًا في تعلم كيفية التطبيق الصحيح لأحكام الشريعة على ما يجد من أمور وأحداث ومصطلحات والاجتهاد في استنباط أحكام شرعية لها بالاجتهاد المنضبط هذا[1].

أما علم الاقتصاد التقليدي أو الوضعي، فيقوم على أساس وسائل قياسية بعينها مثل الموازين المتعارف عليها من موازين سلعية وموازين لقياس درجات الحرارة، ولكن أيضًا يتسع لاستخدام معايير Standards تتمثل في عبارات دقيقة تستخدم لأغراض التقييم للأمور غير الملموسة أو غير القابلة للكيل اليدوي. ومن الأمثلة على تلك **المعايير التقييمية أو الترجيحية** يمكن أن نذكر الآتي على سبيل المثال لا الحصر:

(1) معايير تستخدم لتقييم **السلبيات/ الإيجابيات** أو المفقودات/ الإضافات أو المكاسب والأرباح/ الخسائر أو المحاسن/ المساوئ... وما شابه ذلك. وهناك من المعايير ما يمكن أن تستخدم في حالات بحثية تقوم على دراسات جدوى لمشاريع معينة أو لأنشطة معينة أو لأهداف أو لخطط يؤمل تحقيقها، بل ويمكن أيضًا أن تستخدم بين الخيارات المتاحة أو تحديد البند الاستهلاكي الأفضل بين المتاح والمعروض ... الخ ومن بين تلك المعايير يمكن أن نذكر الآتي:

---

(1) زينب صالح الأشوح، 97، ص 185-186.

أ - التكلفة/ العائد Cost / Benefit (خاصة إذا كانت التكلفة تحتل الاهتمام الأكبر لدى الباحث).

ب- العائد/ التكلفة Benefit / Cost (خاصة إذا كان العائد يحتل الاهتمام الأكبر لدى الباحث).

ج- التكلفة/ التكلفة Cost / Cost (خاصة إذا كان تدنية التكاليف هو المطلوب بغض النظر عن بند العوائد واستهداف تحقيقها).

د - العائد/ العائد Benefit / Benefit (خاصة إذا كان تحقيق أكبر عائد ممكن يمثل الهدف المحوري بغض النظر عن بند التكاليف).

هـ - تكلفة الفرصة البديلة Oportuity / Cost: وتعني قيمة المضحى به في سبيل الاستفادة بالحالي كبديل.

و - القيمة الحالية Present worth ، أو القيمة السنوية Annual worth ، أو القيمة المستقبلية Future worth وكلها تنتمي إلى عائلة معيارية واحدة تسمى القيمة المتعادلة Equivalent worth وهي تستخدم في تحويل جميع التدفقات النقدية عند نقطة أو نقط زمنية معينة باستخدام معدل سنوي للفائدة مثل فرض الضرائب) بما يعادل أعلى معدل للدخل يمكن أن يجذب المستثمر للحصول على قرض ما[1].

(2) كما أن هناك **معدلات ونسب ومؤشرات Rates, Percentages, Indicators** يمكن استخدامها على سبيل الاستدلال Indication أو التعرف القياسي القابل لإدراك الأمور مثل:

أ - معدلات السرعة، المعدلات الضريبية، معدل طبيب/ سكان، معدل الفاعلية، معدل الإنتاجية، معدل النمو الاقتصادي، معدل البطالة، معدل الشفاء...

---

(1) زينب صالح الأشوح، 1994، ص 91-92.

وفي تلك الحالة يجب التنبيه على أن مثل تلك المعدلات Rates or Ratios لا يشترط أن تمثل نسب مئوية. فعلى سبيل المثال قد يشير معدل طبيب/ سكان إلى متوسط عدد السكان الذين يلتزم طبيب واحد بمعالجتهم ويحسب كالتالي:

$$\text{معدل طبيب/ سكان} = \frac{\text{عدد الأطباء المسجلين بوزارة الصحة}}{\text{إجمالي عدد السكان}} \times 10.000 \text{ فرد}$$

ب- أما النسبة المئوية Percentage فهي معدلات مضروبة × 100 بحيث لا تشير إلى أعداد مطلقة لكن إلى أرقام نسبية مئوية مقارنة بأرقام أصلية تتم المقارنة بها.

$$\text{مثل النسبة المئوية للبطالة} = \frac{\text{عدد العاطلين}}{\text{إجمالي المسجلين في القوى العاملة}} \times 100$$

فإذا كان الرقم «35» على سبيل المثال، فذلك يعني إن العاطلين يمثلون 35٪ من إجمالي العاملين بدولة ما. ومن الطبيعي أن الباحث يجب أن يحدد في نطاق بحثه الفترة الزمنية محل الاعتبار بحيث يحصل على بيانات حول إجمالي المسجلين في القوى العاملة، وحول عدد العاطلين عبر سنوات تلك الفترة ثم يقوم باستخراج نسبة هؤلاء العاطلين (المئوية) إلى اجتماع العاملين في القوى العاملة.

(3) هذا، ويمكن أيضًا قياس الدراسات الحديثة على **نظريات علمية أو قوانين أو مبادئ علمية** متخصصة حازت على قبول الأوساط المعنية المتخصصة. بل إن أي بحث علمي يخلو من استخدام النظريات والقوانين والمبادئ العلمية المتخصصة غالبًا ما يرفض بحثه لأنه بني على فراغ أو خواء علمي أساسي. وبناءً على ذلك، فإن الباحث في مجال علمي يجمع بين تخصص علمي بعينه من ناحية، وبين تخصص تشريعي ديني من ناحية أخرى. لابد أن يراعى التقصي الدقيق عن النظريات أو القوانين التي يقع بحثه في دائرتها المعرفية، وكذلك النصوص الدينية والأحكام الشريعة التي يمكن أن تكون ذات صلة بذات المجال، ثم يطبقهما معًا في توليفة متناغمة بتصرف عقلاني تحليلي لا يشذ عما استعان به من أصول علمية تخصصية وضعية (نظريات، قوانين) أو دينية (نصوص، أحكام ...) ومن الأمثلة على ذلك:

أ – نظرية **مالتس في السكان** ونظريات التجاريين والطبيعيين وغيرها من النظريات التي تمت في ذلك المجال المرتبط بالسكان. وترجيح تلك النظريات بالنصوص والأحكام الشرعية المقابلة، ويستخدم كل هذا كأساس نظري وترجيحي وتوضيحي لدراسة يقوم بها الباحث حول مشكلة تزايد السكان (أو انخفاض حجمهم) في دولة ما، مع تقصى العوامل المرتبطة بتلك المشكلة والمقترحات المعروضة لكيفية التعامل معها.

ب– القانون **الحديدي للأجور** الذي يركز على وجود آلية تلقائية لضبط مستويات أجور العاملين، ونظرية ماركس في العمالة التي تعطي للعمالة أولوية مطلقة في حق الاهتمام والرعاية، ويقابل ذلك المنظور الإسلامي الذي يركز على واجبات كل من العامل (بوجوب إتقان عمليه) وصاحب العمل (بوجوب إعطاء العامل أجره وعدم المماطلة في ذلك، مثلما يقول الحديث الشريف «أعط الأجير أجره قبل أن يجف عرقه».

ج– مبدأ «**تساوي بقية العوامل**» الذي يفترض أن كل العوامل لها قيمة تأثيرية بدرجة واحدة على أمر أو على ظاهرة ما ولذلك فعند التحليل لابد من أخذ كل العوامل في الاعتبار والتقدير، بينما يقابل ذلك مبدأ بديل هو «**بقاء العوامل الأخرى على حالها**» الذي يقتضي دراسة تأثيرات بعض العوامل المعتقد في زيادة أهميتها مع تجاهل العوامل الأخرى خاصة ما يتصور بأنها لا علاقة مباشرة لها بالمتغير محل الدراسة والتحليل. أما في المنظور الإسلامي فيقتضي على الباحث أن لا يغفل عن مبدأ «**القدرية**» والمشيئة الإلهية الذي قد يلعب العامل الأكبر في تأثيره على المتغير محل الدراسة.

وفي ذلك لا أنسى أبدًا إنني طلبت من مسئولين كنت أشاركهم في وضع التخطيط الإستراتيجي لمدينة معينة بمصر عدم إطالة أمد عرض التوقعات لأن هناك أقدار تغفل تأثيراتها؛ لكن الجميع أصروا على أن الدراسة المقدمة في أواخر عام 2009 لابد أن تقدم صورة تنبؤية حتى 2050 وفجأة، قامت ثورة 25 يناير 2011 التي

قلبت كل الموازين وأطاحت بتصوراتنا المبنية على عوامل تأثيرية مؤقتة بفترة بدء الدراسة. وأصبح علينا تقديم تصورات أخرى أكثر منطقية تتلاءم والحالة الاقتصادية المستحدثة والمتدهورة بشدة للاقتصاد المصري بعد قيام الثورة المفاجئة وما تبعها من أحداث متشابكة لاحقة تدفع الجميع بدون استثناء داخل مصر وخارجها لإعلان عجزه الكامل عن التنبؤ بالمصير النهائي لمصر ولشعبها بل وللمنطقة المحيطة بأسرها.

رابعًا: الوسائل المحاسبية Accounting Methods

وهي ترتبط عادة بالمساءلة وبالحسابات المالية والسلعية وبطرق تدوينها ومراجعتها، وبتقدير مدى دقتها وضبط ما قد تتضمنها من ثغرات ومن أخطاء، ويمكن التعرف على تلك الوسائل وطبيعتها المتميزة من خلال التعرف أولاً على مفهوم علم المحاسبة ذاته. ومن أهم ما ذكر بهذا الصدد تعريف لجنة المصطلحات التابعة للمعهد الأمريكي للمحاسبين القانونيين AICPA حيث تم النص على أن «المحاسبة هي فن تثبيت وتصنيف وتلخيص بشكل كبير من خلال استخدام تعبير **نقدي** عن صفقات وأحداث التي هي على الأقل جزئيًا ذات **سمة مالية** وتفسير النتائج المترتبة على ذلك» [1]. كما تم تعريفها على أنها «نشاط خدمي وظيفته توفير معلومات **كمية ذات طبيعية مالية بشأن الوحدات الاقتصادية** التي يستهدف منها أن تكون نافعة في صنع القرارات الاقتصادية وفي التوصل إلى خيارات حاسمة بين المسارات البديلة المتاحة»[2].

وبناء على ما سبق، واستنتاجًا لعموم الدراسات المحاسبية يمكن ذكر بعض الأمثلة العامة على وسائل التحليل المحاسبي وهي باختصار شديد:

---

(1) أحمد رياحيي، 2009، ص 81.

(2) المرجع السابق، ص 87.

- القوائم المالية التي تساعد على اتخاذ القرارات المالية.
- المعايير الدولية، ومعايير الإفصاح وغيرها من المعايير ذات الصلة.
- وسائل الإثبات (مثل الجرد والاستدلال على صحة السجلات).
- حسابات وقوائم التكاليف بمختلف أنواعها.
- وسائل الرقابة والمراجعة الداخلية والخارجية.

وبطبيعة الحال، فإن الكتاب الحالي لا يقدم دراسات متعمقة في كيفية التحليل المالي لكل علم على حده، ولا حتى دراسات متعمقة حول وسائل التحليل في حد ذاتها أو بذاتها، لكن المقصود بذلك العرض هو تقديم صورة إجمالية إرشادية للباحث في أي مجال علمي من أجل تحديد الوسائل التحليلية الملائمة لمجال تخصصه الذي قد يختلط مع تخصصات أخرى بحيث يقترن ذلك بزيادة احتمالات رفض بحثه تبريرًا بأنه يتم من منظور تخصصي مخالف لتخصص الواجب الالتزام بالأدوات التحليلية الخاصة به.

ومن أجل تحديد الوسائل المناسبة والصحيحة للتطبيق في دراسة ما، لابد أولاً من تعريف العلم والفرع العلمي التخصصي محل البحث بدقة وذلك من خلال مرجعيات متفق على مصداقية متضمناتها ثم حصر الوسائل المعنية الواردة في نطاق المفاهيم المعروضة وتوضيح الاستخدامات والتطبيقات لكل منها، ثم تحديد ما سيتم اختياره للتطبيق في الدراسة الحالية استرشادًا بما ورد في كتب الأسس والمبادئ العلمية التخصصية (كأدلة نظرية) وكذلك بما ورد في الدراسات بما تمت دراسته في خلال سنوات التعليم الانتظامية، وفترات التدريب الأكثر تعمقًا.

وبعد استعراض بعض الأساليب التحليلية الشائع استخدامها على وجه التعميم فسوف يكون من الأكثر إفادة للباحثين الغوص بتعمق أكثر في أغوار أكثر الأساليب التحليلية شيوعًا في الاستخدام بين المجالات العلمية البحثية المختلفة، وذلك ما نود عرضه فيما يلي من صفحات.

2/1/3 شرح تفصيلي وتطبيقي لبعض الأساليب المختارة

في الجزء السابق مباشرة، كان محور الاهتمام يرتكز على التعريف بهوية الوسائل التحليلية من حيث علاقاتها وصلتها بمسميات اصطلاحية علمية تخصصية محددة، أما في الجزء الحالي، فالاهتمام ينصب على الوسائل والأدوات التحليلية التي يشيع استخدامها في جميع المجالات البحثية بشكل إجمالي وبغض النظر عن التخصصات العلمية التصنيفية، حيث تستخدم تلك الوسائل بشكل أساسي منفرد، أو كوسائل تكميلية وداعمة لأساليب تحليلية أخرى أساسية تخص مجالات معينة مثل الحال في الدراسات العملية في مجالات مثل الطب والهندسة والصيدلة حيث يكون التحليل الأساسي مرتبطًا بطبيعة الدراسة من تشريح أو تجارب ورشية أو معملية لكن ذلك في النهاية دائمًا ما يتطلب إبراز نتائج مثل تلك الأبحاث العملية أو المعملية أو التطبيقية من خلال استخدام وسائل تحليلية وصفية وكمية ملائمة.

أولا: الإحصاء الوصفي Descriptive Statistics

ومن أهم أسباب استخدامه وتطبيقه: تقديم صورة توضيحية رقمية موجودة لوصف البيانات والمعلومات التي تجمعت وتوالدت وتنامت خلال الدراسة؛ بالإضافة إلى الاستدلال على مدى دقة البيانات والمعلومات والنتائج، مع حساب وتقدير الدقة النسبية للقياسات المستخدمة، والمساعدة على التوقعات المستقبلية المحسوبة. كما أنه يفيد في عمليات المسوح الميدانية Surveys وفي القيام بوظائف علمية تحليلية أخرى يمكن الوقوف عليها من خلال التطبيق العملي لكل دراسة علمية جادة.

ويما يلي بعض الأدوات الإحصائية الشائع استخدامها في المجال التحليلي:

(1) قياس النزعة المركزية

أ- حساب المتوسط الحسابي Arithmetic mean or average

وعادة ما يستخدم في حالة وجود مجموعة من القيم المختلفة لوحدات يستهدف

التعرف على الوضع الحالي التقريبي للوحدة الممثلة لتلك القيم، مثل متوسط دخل الفرد الذي يستهدف حسابه من خلال قيم متباينة من دخول الأفراد في نطاق الدراسة المستهدفة.

ويتم حساب المتوسط باتباع الخطوات الآتية:

- تقاس قيمة كل وحدة مفردة بضرب وحدة القياس المستخدمة في تقييمها بعدد مرات حدوث تلك القيمة لذات الوحدة (تكرارها).

وكمثال: إذا كانت الوحدة تتمثل في ش (مثلاً الدخل الشهري لكل العاملين في منشأة ما)، واستهدف حساب متوسط الدخل السنوي للعامل، وكان يعمل في تلك المنشأة عشر أشخاص يحصل كل منهم على دخل شهري مختلف عن الآخر. إذن يتم ضرب الدخل الشهري لكل من هؤلاء العاملين العشرة × 12 شهر.

- يتم جمع كل القيم المختلفة الناتجة في الخطوة السابقة.

(وهي في مثالنا تمثل مجموع حواصل ضرب الدخل الشهري لكل عامل × 12 شهر الممثل لتكرار حصول العامل عليه خلال عام.

- يتم قسمة إجمالي المجموع على المفردات التي يستهدف قياس المتوسط المتعلق بها فيكون الناتج ممثلاً للمتوسط المطلوب حسابه.

وباختصار فإن حساب المتوسط يتم بناءً على المعادلة الإحصائية Formulas التالية:

$$\text{م} = \frac{\text{مج س}}{\text{ن}}$$

حيث «م» تمثل المتوسط الحسابي، «س» تمثل مراكز الفئات والكمية التي تمثل قيمة كل وحدة من المفردات المعنية على حده، «ن» تمثل القيمة الكلية للمفردات.

وبوجه عام، فإن حساب المتوسط الحسابي يتطلب أولاً إعداد جدول تعرض فيه البيانات المرتبطة بالمفردات وبالقيم المناظرة لكل مفردة وبالتكرارات وبحواصل

ضرب كل قيمة مفردة في التكرار الحاصل لها، بحيث يسهل حساب حاصل جمع المفردات وحاصل جمع القيمة المناظرة لإجمال المفردات، تمهيدًا لاستخدامها في التعويض في المعادلة المستخدمة في حساب المتوسط الحسابي. ويمكن توضيح ذلك من خلال الجدول التالي كمثال تصوري.

جدول (1) الوسط الحسابي للدخل السنوي للعاملين بمصنع للحلاوة الطحينية(*)
(بالمائة جنيه مصري)

رقم العامل	1	2	3	4	5	6	7	8	9	10	
الدخل الشهري الذي يحصل عليه	1	1	1	3	3	3	3	8	10	12	
عدد مرات الحصول على الدخل خلال العام	6	6	6	12	12	12	12	12	12	12	
إجمالي الدخل السنوي	6	6	6	36	36	36	36	96	120	144	522

(*) أرقام تصويرية على سبيل المثال

وفي المثال، افترضنا أن هناك عشرة عاملين يحصلون على أجور ومرتبات متفاوتة نتيجة لتأديتهم لوظائف مختلفة في مصنع لتصنيع الحلاوة الطحينة التي تشهر بها مصر بها مصر على الصعيد العالمي أجمع. من بين هؤلاء يوجد ثلاثة عاملين يعملون كعمالة موسمية حصيلتها ستة أشهر وذلك للقيام بأغراض النقل والتسويق والخدمات المتناثرة التي لا تتطلب استمرارية تواجدهم بالمصنع على مدار كل أيام العام وهو في مثالنا الممثلين بأرقام 1، 2، 3. وبطبيعة الحال هم يحصلون على أدنى قيم للمرتبات التي تصل متوسطاتها في الشهر الواحد إلى مائة جنية فقط.

أما العاملين 4، 5، 6، 7 فيقومون بأعمال الإنتاج والتغليف والتعبئة والعرض، وكل منهم يحصل على دخل شهري قيمته 300 جنيه. وبالإضافة إلى تلك العمالة المنتظمة يوجد مدير عام للمصنع يحصل على أعلى راتب شهري وهو رقم «10»

حيث يحصل على 1200 جنيه شهريًا، ويعاونه مدير مساعد يقوم بأعمال الإشراف والمتابعة وهو رقم «9» حيث يحصل على ألف جنيه شهريًا.

وبالتعويض في معادلة المتوسط الحسابي يصبح المتوسط السنوي لأجر العاملين بالمصنع المذكور كالتالي:

$$\text{متوسط الدخل السنوي للعامل} = \frac{\text{إجمالي الدخل السنوي}}{\text{عدد العمال}} = \frac{52200}{10} = 5250 \text{ جنيه مصري}$$

فإذا أردنا التعرف على متوسط الدخل الشهري للعامل يقسم ا لناتج على 12 شهر، أي:

$$\text{متوسط الدخل الشهري للعامل} = \frac{5220}{12} = 435 \text{ جنيه مصري}$$

وبطبيعة الحال فإن ذلك المتوسط لا يمثل المستويات المتباينة للدخول أفضل تمثيل حيث يلاحظ أنه يفوق 4 أضعاف دخل العامل الموسمي وينخفض إلى أقل من نصف الدخل الشهري لأعلى الموظفين درجة، ويربو على الدخل الشهري للفئة العاملة الوسطي بأكثر من ربع قيمة الدخل الشهري لكل من عمالها (والذي قدر بـ 300 جنيه).

وبناءً على ذلك المثال الذي لا يختلف في اتجاهه كثيرًا عن الأحوال المتشابهة نتيجة للتباين في القيم المفردة بحل الحساب، فيجب على الباحث أن يأخذ مثل تلك الوسيلة الحسابية التقديرية بشيء من الحذر وأن لا يعتبر النتيجة الحسابية مؤشرًا كافيًا لتقدير الوضع المستهدف توضيحه وللتعبير عن حقيقة حجمه وقيمته.

بـ الوسيط Median:

ويعتبر من ذات عائلة الوسط الحسابي من دلالة ومؤشر ومغزى، فالوسط يعني الرقم الأوسط Middle Number، وهو يدل على أي القياسات أو المفردات في المجموع الكلي الذي يمكن اعتباره أكثر الأرقام مثالية.

ويتم تحديد الوسط باتباع الخطوات التالية:

- ترتيب القيم أو القياسات أو المفردات ترتيبًا تنازليًا Descending أي يبدأ بالأعلى رقم فالأقل فالأقل حتى ينتهي إلى أصغر قيمة.
- يبدأ الباحث العد الزوجي من كل طرفين (أعلى رقم وأدنى رقم)، ويظل ينتقل في عملية المعدّ الزوجي إلى الطرفين الأدنى في مستوياتهما (الأعلى والأدنى) حتى يصل إلى الرقم الذي يتوسط تلك الفئات المتطرفة المتناظرة فيكون هو الوسيط المعنى. ويمكن أن يتضح من الجدول التالي على سبيل المثال الإرشادي

جدول (2) الوسيط للدخول السنوية بين دخول العاملين بمصنع للحلاوة الطحينية (بالمائة جنيه مصري)

رقم العامل	10	9	8	7	6	5	4	3	2	1
الدخول السنوية وفقا للترتيب التنازلي	144	120	96	36	36	36	36	6	6	6

بالنظر إلى جدول (2) نبدأ أولاً بالمزدوج المتطرف الذي يشمل على أعلى دخل (14400 جنيه) للعامل رقم (10) والأدنى دخل (600 جنيه) للعامل رقم (1)، ثم ننتقل إلى الثنائي المتطرف التالي (120) كأعلى ثاني رقم للعامل رقم (9) و 600 جنيه كأدنى ثاني للعامل رقم (2) ثم ننتقل إلى الثنائي الثالث المتطرف (9600 جنيه، 600 جنيه) للعاملين رقمي (8) و(3) على الترتيب ثم الثنائي الرابع (3600 جنيه، 3600 جنيه) للعاملين رقمي (7 و4 على الترتيب) ثم الثنائي الخامس (3600 جنيه، 3600 جنيه) للعاملين رقمي (6 و5 على الترتيب) وبالمصادفة أن آخر ثنائي وصلنا إليه يحمل ذات الرقم وبالتالي يكون الوسيط هو (360) .

جـ المنوال Mode

وهو المكون الثالث من عائلة قياسات النزعة المركزية وقد يكون الأكثر شيوعًا

في التوزيع بالنسبة لمجموعة من القيم المختلفة، ويتم تحديده بناءً على الأكثر تكرارًا أو الأكثر عددًا من المرات وذلك يتطلب إضافة خانة أخرى في جدول (1) أو جدول (2) يشمل على عدد تكرارات ذات الرقم. وبالنظر إلى جدول (2) يتضح أن رقم «3600» هو الأكثر تكرارًا لأنه تكرر لدى أربعة من العاملين هم 4، 5، 6، 7.

ويلاحظ هنا أن الوسيط يتعادل في قيمته مع المنوال، بينما يبتعد في قيمته كثيرًا عن المتوسط الحسابي (435) وذلك نتيجة للتباين في توزيع الدخول كما أوضحنا من قبل.

(2) قياس التشتت أو الانحراف Measure of Dispersion or variability

في المجموعة السابقة من القياسات، تم إلقاء الضوء على كيفية اختيار أقرب قيمة يمكن أن تمثل مجموعة قيم متباينة بشكل أقرب إلى واقع بقية وقيم المفردات جميعها. لكن برغم ذلك، فقد لوحظ أنه لا يوجد أبدًا ذلك الرقم السحري الذي يمثل كل القيم المتباينة بشكل متقارب أو على نفس الدرجة التي يتقارب بها مع بقية القيم الأخرى. ومن هنا برزت أهمية استخدام مجموعة مقاييس التشتت الحالية وذلك من أجل التعرف على درجات التباين والتشتت بين المتوسط أو الوسط من ناحية وبين المفردات أو القياسات ذات القيم المتباينة من ناحية أخرى، ومن أهم وسائل قياس ذلك التشتت أو الانحراف بين المتوسط المحسوب وبين القيم:

أ- المدى Range

وهو يعبر عن حدود القياسات (أعلاها وأقلها)، ويتم حسابه لاستخدامه للأغراض القياسية على النحو التالي:

- يتم ترتيب مفردات القياسات تنازليًا أو تصاعديًا في قائمة موحدة.
- يكون المدى هو ناتج طرح أدنى رقم بين المفردات المقاسة وبين أعلى رقم فيها.
  أي أن المدى Range = أعلى رقم − أدنى رقم.

- وكلما كانت قيمة المدى صغيرة كلما دل ذلك على تجانس مجموعة المفردات المقاسة، وعلى العكس، فكلما كانت قيمة المدى كبيرة كلما دل ذلك على وجود تشتت وتباين بين قيم المفردات المقاسة وبالتالي عدم تجانس بين قيم تلك المفردات.
- هذا ويلاحظ أنه قد تتقارب قيمة المتوسط لمجموعتين منفصلتين من المفردات أو البيانات ومع هذا تختلف قيمة المدى بين المجوعتين بشكل كبير.
- وبوجه عام فالمدى يعتبر قياس بدائي أو أولي لأنه يتحدد بناء على قيمتين متطرفتين فقط من مجموعة البيانات (وهي أعلى القيم وأدناها)، ومن ثم فهو لا يأخذ في الحسبان التباين والتذبذب بين قيم بقية المفردات بين هذين الرقمين المتطرفين.

بـ الانحراف المعياري Standard Deviation

وهو أفضل من المدى في قياس التشتت والانحراف بين مفردات مجموعة من القياسات أو البيانات ويعتبر قياسيًا مثاليًا في عالم الرياضيات وهو الأكثر استخدامًا لقياس الانحراف بين البيانات أو لقياس مثاليًا في عالم الرياضيات وهو الأكثر استخدامًا لقياس الانحراف بين البيانات أو لقياس مدى تجانس مفردات مجموعة ما من البيانات. فكلما كانت قيمة الانحراف المعياري صغيرة كلما دل ذلك على ضآلة تشتت مفردات المجموعة وبالتالي على زيادة تجانسها والعكس بالعكس أي أن هناك تناسبًا عكسيًا بين قيمة الانحراف المعياري لمفردات مجموعة من البيانات أو القياسات وبين مدى تجانس تلك البيانات أو القياسات.

وبعد التقدم الهائل في برامج الحاسب الآلي المتخصص، أصبح من المتيسر حساب الخطأ المعياري ضمن المستهدف من المؤشرات الإحصائية أو الرياضية الأخرى وذلك بالاستعانة بتلك الأداة الإلكترونية الساحرة. أما في حالة الرغبة في إجراء تلك الحسابات بنفسه، فيمكن للباحث الرجوع إلى كتب الإحصاء التخصصية للتعرف التفصيلي ليس فقط على ما أشرنا إليه سابقًا لكن أيضًا على ما عداها من مصطلحات

إحصائية هامة لاستكمال تحليل البحث العلمي بشكل أفضل وذلك مثل دلالات الاختلافات ومستوى الثقة أو المعنوية Test of Significance والفرض الصفري Null Hypothesis، والخطأ المحتمل، والمنحنى المعتدل إلى غيرها مما يمكن أن يستخدم في قياس ووصف البيانات الكمية، بل واللفظية للبحث.

(3) نظرية الاحتمالات Probabilities Theory

هي نظرية تقوم على ملاحظة ما يطلق عليه في علم الإحصاء بقانون المصادفة Law of chance. وطبقًا لذلك القانون فإن الشخص الذي يلقي العملية المعنية عدة مرات Toss. ستظهر له الصورة نصف عدد المرات التي ألقى فيها العملة بينما يظهر له الوجه الآخر من العملة الذي يحمل أرقامًا في النصف الآخر من مرات الإلقاء العشوائي هذا للعملة.

وتلك النتيجة جاءت نتيجة الخبرة وتكرارات التجارب المماثلة السابقة وليس عن تأكد ويقين بحتمية حدوثها. ومن ذلك نشأ ما يسمى بقانون المصادفة Chance الذي يدلنا على أن الصورة والرقم ستتم مناصفة 50٪. وللتعبير عن ذلك بطريقة أخرى يمكن القول بأن احتمال ظهور الصورة هو «واحد». من بين محاولتين أو أن احتمال ظهور الصورة يمثل نصف (½) أي أن نصف عدد مرات الإلقاء ستظهر فيها الصورة.

ومن استخدامات نظرية الاحتمالات أو قوانين «الصدفة» ما يتم في شركات التأمين حيث تضم هذه الشركات سجلات بحالات الوفاة من أجل اكتشاف الأعمار المختلفة التي يموت عندها الناس، ثم يقوم المختصون من الباحثين العاملين بتلك الشركات بحساب الفرص النسبية للوفاة في مختلف مستويات الأعمار، حيث يساعد ذلك في حساب وتحديد أقساط التأمين التي يجب تحصيلها من العميل، كما يستعين المعلم بتلك النظرية من أجل تصنيف طلابه من ناحية الكفاءة الاستيعابية بحيث يستعين بتكرارات نتائج كل طالب ليحدد مستواه إن كان متوسط Average أو إن

كان متفوقًا أو كان ضعيف المستوى على وجه العموم. ويساعده في ذلك في النهاية على التعرف على المستوى التحصيلي والاستيعابي العام لمجموع الطلاب مما يمكنه من وضع أسئلة امتحان معتدلة تتلائم مع جميع المستويات في آن واحد.

لكننا نتحفظ كثيرًا على نتائج تطبيق تلك النظرية لأنها تقوم على قانون «الصدفة» بينما علمنا الله تعالى أنه لا توجد صدف أبدًا ولكن «كل شيء بقدر»، كما أنها تقوم على تخمينات شخصية غير موضوعية، وترتبط أيضًا بلعبة القمار gambling كفكرة أساسية للتفكير في إنشائها حيث كان من الأمثلة التي جاءت لتوضيح آلية عملها، مما ذكره العالم «دي ميير» من أن هناك رجلين يلعبان بالزهر Dice، وكل منهما يرهن على كمية من المال، والرابح هو الشخص الذي يكسب ثلاث مرات من خمس رميات Throws، ولا نريد أن نستكمل المثال لرفضنا لمبدأ ارتباطه بمثال يتناول نشاطًا محرمًا بوضوح في الدين الإسلامي الحنيف.

ثانيًا: أدوات وصفية تحليلية Descriptive analytical tools

وقبل أن نبدأ حديثنا عن المحور الحالي، نود أن ننوه إلى أن ما عرض من عناوين فرعية إنما يتم بهدف تيسير عملية التتبع والمتابعة المتسلسلة من قبل القارئ. لكن واقع الأمر يؤكد على أن هناك تداخلاً كبيرًا بين الوظائف التي تقوم بها الأساليب المعروضة، حتى يمكن أن يستخدم بعض الأساليب وتحقيق مجموعة متكاملة من النتائج والأهداف الوصفية والتصنيفية والتحليلية للبيانات وللمعلومات كلها، معًا. لكن نؤكد أيضًا على أن التصنيف للأساليب المعنية لم يتم استخدامه بشكل عشوائي بحت، لكن تم استخدامه بناءً على تطبيق معيار «الأكثر شيوعًا في الاستخدامات والتطبيقات في الدراسات العلمية المختلفة».

وبعد قيامنا بالتنويه اللازم بهذا الصدد، نبدأ بعرض بعض الأساليب التي يشيع استخدامها في التطبيقات التحليلية المختلفة، في الغالبية العظمى من الدراسات خاصة ذات الصبغة الكمية أو القياسية منها.

(1) المتغيرات Variables

عندما نتحدث عن المتغيرات لابد أن ندرك أننا نتناول أهم لبنة لا يستطيع أي باحث علمي أن يستغنى عنها ليس فقط في ما سوف يعرضه أو يجريه من بيانات أو معلومات أو نتائج أو عمليات تحليلية، ولكن حتى فيما يجب أن يتضمنه بحثه من مكونات وصفية بحتة فالمتغيرات عادة ما تستخدم كصورة كمية للتعبير عن مفردات كمية أو كيفية بحيث تأخذ شكل متغيرات رقمية Numerical (يتم قياسها بأرقام قياسية حقيقية) أو متغيرات صورية Dummy (تعطي لها أرقام تصنيفية أو تمييزية من أجل استكمال عمليات تحليلية تشملها) وهي تستخدم لأغراض أخرى واسعة غير قياس المفردات أو البيانات مثل تكوين المعادلات التي تمثل روابط أو علاقات ذات طبيعية خاصة (كما سيتم التوضيح فيما بعد). ولنتعرف معًا على المزيد المفيد حول تلك الأداة المفتاحية الهامة وذلك من خلال العرض التالي:

أ - **المتغير الرقمي Numerical variable** : هو الذي يأخذ أرقامًا حقيقة قياسية تعبر عنه.

ب- **المتغير الصوري Dummy Variable**: هو متغير صوري يمثل بيانًا وصفيًا أو نوعيًا، وذلك من أجل المقدرة على استخدام تلك البيانات النوعية غير القابلة للقياس الرقمي الفعلي، يتم التعبير عنها بما يطلق عليه بالمتغير الصوري، حيث يعطي لذلك المتغير رقمين فقط 1 & 0 بحيث يعطي رقم 1 للفئة أو الجزئية داخل ذلك المتغير التي يتم التركيز على تحليلها، ويعطي رقم صفر 0 لبقية مكونات الفعل أو المتغير الصوري.

فعلى سبيل المثال إذا كان لدينا بيان «النوع» ونريد إدخاله في تحليل كمي، نحول ذلك البيان إلى رقم صوري Dummy يأخذ رقم 1 للذكور ورقم صفر 0 للإناث أو العكس إذا ما كانت فئة الإناث هي التي ينصب الاهتمام عليها والتحليل القياسي الجاري.

وكمثال آخر إذا أردنا تحليل بيان «الحالة التعليمية» لفئة ما. فيمكن أن يقاس بمتغير رقمي يتمثل في عدد السنوات التي تم قضائها في المراحل التعليمية المختلفة. لكن يعيب تلك الطريقة أن البعض قد يكونوا قد قضوا سنوات أكثر من الآخرين لأنهم رسبوا كثيرًا فزادت بذلك سنوات الدراسة لهم ومن ثم، فهم في واقع الأمر في مرتبة علمية أدنى من غيرهم الذين حصلوا على ذات القدر من التعليم في عدد أقل من السنوات. وقد تكون فئة مماثلة في عدد سنوات الدراسة لكنهم متفوقين وأكثر تميزًا من الجميع لأن العدد الأكبر من سنوات الدراسة بالنسبة لهم كانوا قد قضوه في إتمام مراحل متقدمة من الدراسات العليا أو الحصول على درجات علمية متميزة مثل الماجستير والدكتوراه والارتقاء المهني إلى أستاذ مساعد ثم أستاذ .... وهكذا.

ومن أجل تلافي تلك المشكلة التي قد تتسبب في نتائج تحليلية مضللة، يمكن أن تقاس الحالة التعليمية بمتغير صوري يتكون من مراحل التعليم المختلفة (إلزامي، إعدادي، ثانوي، جامعي، أعلى مثلاُ) بحيث تأخذ إحدى تلك المراحل محل التركيز والاهتمام رقم (1) ويعطي رقم صفر 0 لبقية المراحل. ومن أجل تحقيق مزيد من الدقة في النتائج، يمكن عمل تباديل وتوافيق بحيث يعطي لكل مرحلة رقم (1) والبقية صفر مكونين بذلك عدة متغيرات صورية فرعية لذات البيان (وهو في مثالنًا الحالة التعليمية) وبحيث يتم استخدام متغير صوري فرعي كل مرة ضمن المنظومة التحليلية الأخرى بشكل منفصل، وبذلك يكون الباحث قد طبق نظرية الاحتمالات والتباديل والتوافيق حتى يصل إلى أفضل النتائج وأكثرها واقعية وتمثيلاً للحقائق.

ج- **المتغيرات الداخلية Endoganous**: هي تلك المتغيرات التي تؤثر في النموذج وتتأثر به كذلك.

د - **المتغيرات الخارجية Exogenous**: هي تلك المتغيرات التي تؤثر في النموذج ولا تتأثر به.

هـ - **المتغير التابع Dependent**: هو الذي يتأثر بالمتغيرات المستقلة ويتغير وفقًا لتغيراتها، ويفسر ما يحدث له من تغييرات كنتيجة لتأثيرات المتغيرات المستقلة.. وهو لا يؤثر في تلك المتغيرات. وعادة ما يوجد متغير تابع واحد فقط في كل علاقة دالية واحدة ويدون دائمًا في الجانب الأيمن من المعادلة.

و - **المتغيرات المستقلة Independent أو التفسيرية explanatory**: هي متغير واحد أو أكثر تؤثر في المتغير التابع ولا تتأثر به، وتفسير ما يحدث للمتغير التابع من تغيرات في قيمه أو اتجاهاته، ولكن لا يمكن أن يفسر المتغير التابع بقيمه المختلفة ما يحدث لها من تغييرات قيمية أو اتجاهية.

وإذا كانت هناك مجموعة من المتغيرات المستقلة تؤخذ معًا في معادلة دالية واحدة، فيجب أن تكون مستقلة أيضًا بعضها عن البعض، بمعنى أنه يجب أن لا توجد تأثيرات من أحد تلك المتغيرات المستقلة على أحد أو بقية المتغيرات المستقلة الأخرى محل التحليل، ولا يجب أن توجد تأثيرات متداخلة فيما بينها لأن ذلك يسلب جزءًا من خاصية الاستقلالية Independency التامة التي تتصف بها تلك النوعية من المتغيرات، مما يتسبب في نشوء مشكلة تعددية التأثيرات المتداخلة Multicolinearity التي تتسبب في الخروج بنتائج تحليلية مضللة للعلاقات محل الدراسة.

ز - **المتغيرات الساكنة Static**: هي التي تأخذ قيم لحظية، ولا تحتوي على عنصر الزمن، وعادة ما تستخدم لدواعي تبسيط عمليات التحليل أو تدوين تصورات مبدئية عن أمر ما، أو لوصف مشكلة ساكنة تحتاج لتحليل متعمق مثل تحديد المعايير الإدارية المثلي لنظام ما.

ح- **المتغيرات الحركية Dynamic**: هي التي يلعب الزمن دورًا أساسيًا في تحديد قيمها ومساراتها، بحيث تأخذ قيمًا مختلفة باختلاف الوحدات الزمنية أيا ما كانت قيم تلك الوحدات (أيام، شهر، سنوات، عصور، ...) وعادة ما تنطوي

على نتائج أكثر واقعية ومنطقية مقارنة باستخدام المتغيرات الساكنة، وإن كان يصحب عمليات تحليلها مشاكل وتعقيدات أكثر مثلما يحدث في وصف المنهاج التحليلي أو الإجراءات المرتبطة بها.

وتجدر الإشارة إلى أن كثيرًا من المتغيرات الحركية المستخدمة في نماذج أو عمليات تحليلية ما، لابد من أجل التوصل إلى تحديدها الصحيح أن تسبقها دراسات لمتغيرات ساكنة بحيث يمكن تعديل تلك الدراسات التمهيدية للوصول إلى أفضل مجموعة من المتغيرات الحركية الملائمة للتحليل وللدراسة القائمة على عنصر الزمن.

(2) العلاقات Relationships، الدوال Functions:

بعد أن تعرفنا على المتغيرات التي أشرنا إلى أنها تمثل لبنة أساسية من لبنات أية بناء وصفي أو تحليلي، حان الوقت لنوضح بتفصيل أكثر لماذا ذكرنا ذلك حين نتعرف الآن على أهم هيكل وصفي تحليلي يعتبر بحق قاسمًا أعظمًا مشتركًا كنواة أساسية متكاملة لأية دراسة علمية منطقية مقبولة خاصة إذا ما كانت تتم بهدف البحث عن بعض التأثيرات المستهدفة أو عن العوامل الأكثر تأثيرًا في ظاهرة ما وفي وضع ما أو ما شابه ذلك. ذلك الهيكل الوضعي التحليلي المركزي يتمثل في نوعه الأساسي الذي يطلق عليه علاقة (دالية) أو يطلق عليه مصطلح فردي المسمى وهو دالة Function.

أ- مفهوم الدالة واستخدامها

والدالة هي علاقة سببية بين متغير تابع (يتمثل في الظاهرة أو الوضع محل التقصي والدراسة) وبين متغيرًا آخر مستقل يفترض أن له أولهم تأثيرًا معنويًا على قيم ومسارات المتغير التابع بدرجة أو بأخرى، بالسالب أو بالموجب.

ويمكن استخدام العلاقة الدالية في بداية بحث علمي ما باعتباره «فرضية» تتطلب التحقق من صحتها أو من خطئها، ويمكن أن تعرض في نهاية دراسة تطبيقية

ما وتعتبر في تلك الحالة نتيجة من نتائج التحليل، بحيث تفيد الدراسات المماثلة التالية في إجراء دراسات تحليلية تصحيحية من حيت انتهت الدراسة التي عرضت الدالة كنتيجة، وذلك إذا ثبت عدم صحتها أو عدم تطابقها مع فرضية الدراسة المعروضة بالمقدمة. وإن ثبتت صحة المعادلة، يمكن أن تفيد في التنبؤات القياسية ذات الصلة باستخدام المعادلة المقبولة والمنجزة بنجاح.

ب- بعض أنواع الدوال الأكثر شيوعًا في التطبيق

والدالة إما أن تكون بسيطة بحيث تمثل العلاقة بين متغير تابع، ومتغير مستقل واحد يفترض أنه المتغير التفسيري للمتغير التابع وفي تلك الحالة تأخذ الشكل التالي:

س = د (ص)

حيث «س» ترمز للمتغير التابع، «د» ترمز للدالة وتعني وجود علاقة بين المتغير التابع «س» وبين «ص» الذي يرمز إلى المتغير المستقل.

وغالبًا ما يؤخذ في الاعتبارات التحليلية أكثر من متغير مستقل وبالتالي تأخذ الدالة الشكل التالي:

س = د (ص، ل، م، .... ن)

وفي تلك الحالة نجد أن المتغيرات المستقلة متعددة وهي هنا ص، ل، م، ...، ن حيث أن كل رمز يشير إلى متغير قياسي مستقل ما.

ومن الأمثلة على مثل تلك الدوال دالة الطلب التي يمثل المتغير التابع فيها «الكمية المطلوبة من سلعة ما»، بينما تتمثل المتغيرات المستقلة في سعر ذات السلعة، وأسعار السلع الأخرى، والدخل والأذواق.

(3) النماذج Models:

النموذج التحليلي أو العلمي هو هيكل تحليلي كمي مبسط تعرض من خلاله

مجموعة من المتغيرات لبيان ظاهرة معينة من خلال دوال معينة توضح بعض العلاقات السببية بشكل علمي مبني على ضوابط علمية محددة Formulae. والنموذج يمثل بمكوناته المعروضة ملخصًا بالغ الاختصار لحقائق معينة أو لنتائج تحليلية معينة بحيث يعبر عنه بصورة رياضية يمكن أن تستخدم في قياس متغيرات مماثلة أو قابلة لذات القياس الذي يتضمنه ذلك النموذج. وهو يأخذ أشكالاً وقياسات وتطبيقات مختلفة باختلاف ركائز بنائه أو تطبيق المعد سلفًا من بعض أنواعه.

والنموذج يقوم على خليط من المفاهيم والوسائل والأدوات الرياضية والإحصائية والاحتمالات ونظريات القرارات وغيرها من النظريات والقوانين العلمية المتخصصة ومن الأمثلة على مثل تلك النماذج نذكر نماذج التوازن Equilibrium التي تهدف إلى محاولة تفهم وتحديد العلاقة بين الاستهلاك والإنتاج والأسعار في اقتصاد ما تسوده المنافسة، أو بهدف المساهمة في تحديد المشاكل الاقتصادية والإدارية وغيرها والمساهمة في كيفية المعالجة لتلك المشاكل.

وفيما يلي بعض الأمثلة الوجيزة على أنواع النماذج الرياضية الإحصائية التي تمثل ركائز في دنيا النمذجة العلمية:

أ - **النماذج الساكنة Static Models**: هي التي تقوم على (أو يستخدم في بنائها) متغيرات ساكنة لا تأخذ عنصر الزمن في الاعتبار. وهي تبني وتستخدم من أجل التعامل مع المشاكل والمسائل الساكنة التي تتطلب معالجات أكثر عمقًا في خلال لحظة زمنية معينة.

ب- **النماذج الحركية Dynamic Models**: هي التي تقوم على (أو يستخدم في بنائها) متغيرات حركية تشمل قيمًا مختلفة عبر فترات زمنية معينة بحيث يؤخذ عنصر الزمن في تحديد قيمه المتعددة. وذلك النوع من النماذج يرتكز على دراسة تطور أو تغير أو مدى استمرارية ظاهرة ما عبر فترة زمنية أو تاريخية معينة ومن ثم فهو صالح للاستخدام في عمليات التنبؤ العلمي المستقبلي لما

يمكن أن يحدث في زمن يلي وقت انتهاء الدراسة التحليلية التي عرضت مثل ذلك النوع من النماذج بين طياتها أو كنتيجة مستهدفة بصفة أساسية من إجرائها وتنفيذها. وذلك مثل النماذج التي تتناول نظامًا اقتصاديًا معينًا مثل النظام الإسلامي، أو تلك التي تتناول نظام علاجي لمرض وبائي منتشر، أو تلك التي تتناول نموذجًا معياريًا إسلاميًا أو نموذجًا معماريًا موجهًا لفئة معينة من طوائف الشعب وفي ذلك يحضرنا النموذج الخاص بالمهندس عالمي الصيت الراحل «حسن فتحي» الذي قدم نموذجًا بالغ التميز حول «عمارة الفقراء» والذي يمكن تطبيقه في أية مرحلة زمنية تختلف عن زمن نسج وعرض وتطبيق ذلك النموذج القديم الحديث.

ج- **النماذج المرتبطة بالقرارات الإدارية وغيرها Management Decisions Models**: وهي عادة ما ترتبط في عمليات بنائها بنظرية علمية متخصصة يطلق عليها نظرية المباريات Theory of Games.

د - **النماذج التعظيمية Maximization Models**: وهي تقوم على دوال تستهدف إلى تعظيم شيء ما مثل تعظيم الإنتاج أو تعظيم درجة الفعالية ويطلق على وظيفة معظمة الدوال Functions Maximization. ومثل تلك النماذج تتطلب وضع مجموعة من القيود، وتستخدم عادة البرامج الخطية Linear programming، والبرامج الحركية Dynamic programming.

هـ- **نماذج تعاني من عدم التأكد Uncertainty**: وهي تتطلب استخدام بعض الطرق الإحصائية المساعدة على التعامل الأفضل مع ذلك التصور التحليلي الهام.

و - **نماذج تحديدية Deterministic**: وهي من الناحية الفنية أبسط وأسهل من ناحية المعالجة الرياضية.

والواقع أن استخدام النماذج في البحث العلمي يتطلب علمًا وخبرة وتدريبًا

ومهارة على خصوصية أكثر دقة لما تكتنفه من تعقيدات وقيود ومحددات ومشاكل. لذا، يجب على الباحث الحذر من استخدام ذلك الأسلوب التحليلي ما لم يكن يتمتع بالقدرات وبالإمكانيات الكافية.

وبوجه عام يمكن القول بأن تطبيق دراسة لنموذج سبق إعداده والاتفاق العلمي عليه هو أمر أكثر يسرًا من عملية بناء نموذج كامل. ففي الحالة الأخيرة لابد من دراسة مكثفة ومتعمقة للعلوم الرياضية والإحصائية والكمية والقياسية بالإضافة إلى دراسة المواضيع المتخصصة المتعارف عليها. وعلى الراغب في تطبيق ذلك الأسلوب التحليلي المعقد أن يظل على صلة واستعانة دائمة بمراجع تشمل على أساسيات النمذجة ومتعلقاتها مع التبحر بين الدراسات المتخصصة والمناظرة التي استعانت أو قامت بتطبيق أساليب النماذج وقامت بعرضها بلغة نقاشية وتحليلية وعلمية تتلاءم مع مجال التخصص العلمي محل أو موضع الدراسة كالاقتصاد أو الصحة أو الهندسة ... الخ.

(4) الارتباط البسيط Simple correlation الابن التحليلي المدلل:

إنه بحق يمثل أكثر الأساليب التحليلية المستخدمة في سائر الدراسات العلمية بكل تخصصاتها واهتماماتها لأنه يتميز ببساطة إجراءاته وتيسر عمليات تنفيذه، ولا يتطلب تعقيدات أو مشاكل في توفير البيانات التي يقوم عليها ولا يتطلب أية مهارة كبيرة في فهم ما يعرضه من نتائج رقمية.

فالارتباط هو أسلوب تحليلي Analytical Statistical Technique يقوم على توضيح العلاقة بين متغير تابع وبين متغير واحد (فقط) مستقل وتوضيح درجة التأثيرات التي يحدثها المتغير المستقل عليها. وأهم ما يجب توضيحه في هذا الصدد

أ - **شكل المعادلة المعبرة عن الارتباط:** هي المعادلة الدالية البسيطة التي تشمل على المتغير التابع في الجهة اليمنى، والمتغير المستقل في الجهة اليسرى. وفي بداية التحليل يرفق بالمتغير المستقل حرف «د» للإشارة إلى وجود علاقة (دالية)

تربط بين ذلك المتغير المستقل وبين المتغير التابع الذي يوجد في الجهة اليمنى من المعادلة الذي يفصل بين الجهتين علامة «=».

وبناءً عليه فإن الشكل الأساسي للدالة الممثلة على علاقة ارتباط بسيط بين متغيرين أحدهما تابع «س» مثلاً والآخر متغير مستقل «ص» مثلاً يكون كالتالي:

س = د ص

أما بعد إدخال البيانات (الممثلة للقيم المختلفة للمتغير المستقل «ص»، وقيم «س» المناظرة) في حيز تحليل معامل الارتباط بين المتغيرين المذكورين، تصبح معادلة الارتباط على الشكل التالي:

س = أ + (م × ص) + و

حيث س هي قيمة المتغير التابع المجهولة التي سوف تحسب قيمتها بالتعويض في الجانب الأيسر من المعادلة. وعلى الجانب الأيسر يعبر «أ» عن قيمة الثابت بالمعادلة وهو يتمثل برقم محدد، و «م» يعبر عن معامل يشير إلى قيمة التغيير Change الذي سوف يحدث في قيم المتغير التابع «س» حيث تتغير قيمة المتغير المستقل «ص» بوحدة واحدة. أما الرمز «و» فيمثل ما يسمى بالمتبقيات Residuals الذي يشير إلى قيمة العوامل الأخرى التي لم تؤخذ في الاعتبار أثناء العملية التحليلية والتي يمكن أن تؤثر في مصداقية النتائج المبنية على النموذج الدالي المبني على الارتباط البسيط.

ب- دلالات القيم المختلفة لمعامل الارتباط Coefficient of Correlation

- إذا كان معامل الارتباط «1» فذلك يدل على وجود ارتباط معنوي كامل بين المتغير التابع والمتغير المستقل.
- إذا كانت الإشارة المرتبطة بمعامل الارتباط «موجبة» فذلك يدل على أنه كلما تغيرت قيم المتغير المستقل، تغيرت قيم المتغير التابع في نفس الاتجاه. فمثلاً إذا كنت قيم المتغير المستقل تتزايد، فإن قيم المتغير التابع تتزايد تبعًا لذلك، والعكس بالعكس صحيح.

- إذا كانت الإشارة المرتبطة بمعامل الارتباط «سلبية» فذلك يدل على أن التغيرات في قيم المتغير التابع ستأخذ اتجاهًا معاكسًا لاتجاه التغير في قيم المتغير المستقل، فمثلاً إذا زادت قيم المتغير المستقل، فسوف تنخفض قيم المتغير التابع والعكس بالعكس.

- إذن لابد عند تفسير نتائج الارتباط بين متغير مستقل ومتغير تابع الذي يشير إلى «تأثير» المتغير المستقل على المتغير التابع أن يؤخذ في الاعتبار أمرين جوهريين:

• قيمة معامل الارتباط: التي تشير إلى وجود ارتباط بين عاملين أو متغيرين من عدمه والي درجة ذلك الارتباط إن ثبت وجوده.

• الإشارة المقترنة بمعامل الارتباط: فإذا كانت موجبة فهي تدل على أن العلاقة بين المتغيرين طردية وإن كانت الإشارة سلبية فهي تدل على أن العلاقة بين المتغيرين عكسية.

• وبكلمات أخرى يمكن القول بأن معامل الارتباط يفيد إجمالاً في التعرف على ما إذا كان هناك عامل آخر له علاقة بالعامل أو المتغير محل الدراسة أم لا. فإذا كانت قيمة معامل الارتباط (صفر) فهذا يدل على أن الأشياء التي تتم المقارنة بينها ليست مرتبطة بعضها ببعض وعلى أن المتغير محل الدراسة لا يتأثر في قيمه بالعوامل الأخرى وبالتالي فهو مستقل وغير تابع. أما إذا أخذ معامل الارتباط أية قيمة أكبر غير الصفر (سلبية كانت أم إيجابية) فذلك يعني أن هناك متغيرًا يؤثر في المتغير محل الدراسة بشكل عكسي (إذا كانت النتيجة سالبة الإشارة) أو طردي (إن كانت الإشارة موجبة) وكلما اقتربت القيمة من «1» صحيح كان الارتباط بين المتغيرين قوي، والعكس.

- تقديم صورة مبسطة حول وجود علاقة تأثيرية من قبل المتغير المستقل على المتغير التابع من عدمها.

- تقديم صورة مبدئية تمهيدية لطبيعة العلاقة بين متغيرين أحدهما تابع والآخر

مستقل تسمح بإجراء عمليات ارتباط تبادلية لهذين المتغيرين بحيث يتم الوقوف على حقيقة المتغير التابع منهما وعلى المتغير المستقل في حقيقة الأمر.

- تقديم صورة أولية لعلاقات مسببة يتم استخدامها كأسس مرشدة أولية من أجل التقدم بصورة تحليلية أكثر موضوعية ومصداقية تتمثل فيما يسمى بالارتباط المتعدد Multiple correlation حيث يتم بموجبه دراسة تأثيرات أكثر من متغير مستقل على المتغير التابع، وكلما زاد عدد المتغيرات المستقلة المأخوذة في الاعتبار التحليلي، كلما ساعد ذلك على الوقوف على نتائج محايدة وأقل تحيزًا لمتغير دون الآخر.

(5) الانحدار المتعدد Multiple Regression Analysis

هو بحق «ملك» الأساليب التحليلية حيث يشيع استخدامه وتطبيقه كأسلوب تحليلي متكامل أكثر قبولاً واعتبارًا في استنباط النتائج التي تؤخذ بمزيد من الجدية والاعتبار، ويمكن تطبيقه وحده أو بوسائل تحليلية مساعدة مثل أسلوب الارتباط البسيط سالف الذكر، أو تطبيقه ضمن منظومة تحليلية متكاملة مثلما يحدث في حالة النمذجة التي تقوم على بناء أو تعديل أو تطبيق نماذج رياضية إحصائية ما.

وتحليل الانحدار المتعدد أو ما يطلق عليه أحيانًا تحليل الانكفاء، هو أسلوب إحصائي يتم بمقتضاه التنبؤ بمتوسط متغير عشوائي أو عدة متغيرات عشوائية اعتمادًا على قيم وقياسات متغيرات عشوائية أخرى، وذلك الأسلوب التحليلي له عدة أنواع منها الانحدار الخطي (وهو الأكثر شيوعًا في التطبيق نتيجة للسهولة النسبية لاستخدامه ولتطبيقه)، والانحدار اللوجيستي وانحدار بواسون. ومن أهم استخداماته ووظائفه المستهدف تحقيقها:

- التنبؤ بمتغير مستقل من خلال التعرف على قيم معطاة للمتغير التابع أو العكس.
- تحديد نسبة إسهام كل متغير مستقل في التأثير على تحديد قيم المتغير التابع وعلى المتغيرات التي تطرأ عليها وذلك من خلال نموذج خطي Linear تعرف معادلته باسم معادلة الانحدار الخطي المتعدد فيما يلي:

أ - **تحديد المتغير التابع Dependent Variable وكيفية قياسه**: وهو يمثل عادة جوهر الموضوع التحليلي المستهدف بحثه والتعرف عليه أو على أمور علمية تتعلق به، وتحديد كيفية قياسه بما يتلاءم مع إجراء عملية تحليلية صحيحة باستخدام أسلوب الانحدار المتعدد وباستخدام البرنامج الحاسوبي المناسب لذلك.

فعلى سبيل المثال إن كانت هناك دراسة ما تهدف إلى بحث الآثار الاقتصادية والاجتماعية للأمهات على الحالة الغذائية للأطفال دون سن الدراسة في مصر. فإن المتغير التابع في تلك الحالة يتمثل في الحالة الغذائية للأطفال المعنيين. لكن الاكتفاء بذكر هذا لا يقبل كدراسة علمية دقيقة لأن الحالة الغذائية يمكن أن تترجم إلى معان كثر، ويمكن أن تشمل على حالات فرعية متباينة تتسبب في حدوث خلط وعدم دقة في تحديد المتغير التابع وهو ما يعني استخدام بيانات تحليلية غير دقيقة مما يزيد من احتمالات خطأ النتائج النهائية وبالتالي عدم نجاح الدراسة في تحقيق هدفها على النحو العلمي الصحيح.

إذن، فعلى الباحث في تلك الحالة أن يغطي كل الدراسات المتخصصة في مجالات التغذية والصحة العامة بشكل واف، وأن يستخرج المعايير الشائع استخدامها في قياس الحالات الغذائية للفئة المعنية ويستعرض طرق قياسها، ودلالات ما تشمل عليه من مؤشرات ثم تتم المفاضلة فيما بينها على أسس موضوعية تتلاءم مع طبيعة البحث (إن كان طبي متخصص أو اجتماعي غير متخصص مثلاً) وتتلاءم مع خطة استقطاب البيانات اللازمة بالإضافة إلى أمور أخرى يتم التعرف عليها من الدراسات المماثلة السابقة للبحث الجاري إعداده. وفي نهاية الأمر يتم تحديد المتغير التابع بشكله النهائي الذي يمثل الحالة الغذائية المعنية. وليكن في مثالنا الحالي الوزن/ الطول (أي وزن الحالة مقارنة بطولها).

وبعد تحديد المتغير التابع لابد من تحديد كيفية قياسه وعرضه في شكل رقمي لإمكانية تطبيق الإجراءات التحليلية عليه. وفي مثالنا الحالي يتم قياس الحالة

الغذائية لكل مفردة بقسمة وزنها (بالكيلوجرام أو بالرطل مثلاً) على الطول (بالسنتيمترات أو بالبوصات مثلاً).

ب- **تحديد المتغيرات المستقلة Independent variables وكيفية قياسها**: وهي تلك المتغيرات التي يتوقع أن يكون لها أكبر تأثير على تحديد وتغيير قيم المتغير التابع، والتي تقرر إدخالها في معادلة الانحدار الخطي المتعدد.

ويتم تحديد كل متغير مستقل على نحو ما تم شرحه توا فيما يتعلق بالمتغير التابع، مع مراعاة تحديد صور المتغيرات المستقلة التي تتلاءم مع المتغير التابع المختار. ذلك، مع وجوب مراعاة أن تكون التأثيرات فيما بين مجموعة المتغيرات المستقلة صفرية أو في أقل قيم لها تجنبًا لمشكلة تداخل مثل تلك التأثيرات بين المتغيرات المستقلة على تأثير كل منها على قيم المتغير التابع مما يتسبب في الخروج بنتائج مضللة أو غير دقيقة أو غير مفسرة بشكل واضح للتأثيرات محل البحث والاهتمام. فعلى سبيل المثال، عندما تتحدث المتغيرات المستقلة الممثلة للحالة الاقتصادية للأمهات يجب عدم اختيار «متوسط دخلها» وفي نفس الوقت «متوسط ما تنفقه على الطعام». ذلك لأن ما ينفق على الطعام هو جزء من دخلها - وبالتالي فإن المنفق على الطعام هنا يمثل متغيرًا تابعًا لمتوسط دخل الأم وليس مستقلاً بذاته، فإذا ما وضع الاثنان في الجانب الأيسر الخاص بالمتغيرات المستقلة معًا، فسوف يحدث تداخل كبير في تأثيراتهما على الحالة الغذائية للطفل ويصبح من الصعب معه التعرف على تأثير كل من المتغيرين على حده على المتغير التابع المذكور.

وبالإضافة إلى ما سبق، يجب أيضًا تحديد طبيعة المتغير إذا ما كان رقميًا numeric فكيف يقاس وما هي وحدة القياس التي سوف تستخدم في قياسه. أما إذا كان اسميًا String فيجب تحديد صورة المتغير الذي سوف يمثله كمتغير صوري Dummy بأخذ رقمين تصنيفين فقط هما واحد وصفر، وذلك على نحو ما شرحناه تفصيلاً في خلال حديثنا حول المتغيرات.

ج- **تحديد اختبارات الفروض والمعنوية The Null hypothesis & Significance**:
ستتم الاستعانة بها من أجل التحقق من مدى حيادية ومصداقية نتائج التحليل والبيانات التي تم استخدامها من أجل إتمام عمليات التحليل المستهدفة.

د - **تحديد البرنامج الحاسوبي الملائم للتحليل وحتى الآن هو «الحزم الإحصائية للعلوم الاجتماعية»**. فمع التطور المعاصر الضخم في مجال استخدامات الحاسوب الآلي، أصبح استخدامه من أجل تنفيذ العمليات التحليلية المستهدفة هو القرار الحكيم بشرط استخدام البرنامج الملائم لطبيعة البيانات والتحليل المستهدف من ناحية، وبشرط التعرف على كيفية إدخال البيانات الخام محل التحليل بوضعها الصحيح من ناحية أخرى، مع التعرف الكافي على كيفية تشغيل البرنامج الحاسوبي وعلى كيفية قراءة وفهم ما تتوالد خلاله من نتائج رمزية ورقمية تتطلب استعادة العرض والتوضيح اللفظي لها. وذلك على نحو ما سيتم عرضه بشكل أكثر تفصيلاً في جزء لاحق وشيك.

هـ - **استخراج النتائج الرقمية والرمزية وعرضها في صورة معادلة الانحدار المتعدد:**
مع إلحاق عرض تلك المعادلة مع عرض نتائج الاختبارات الإحصائية ذات الصلة والدلالة بحالها الرقمي المستخرج كأن تكون المعادلة المستنبطة من مخرجات التحليل الحاسوبي مثلاً:

و ع = 1.91 + 0.6 − 0.03 ح

حيث R^{-2} = ... وكا تربيع ^{2}(كا) = ... ، مستوى المعنوية = ...

وما من شك في أن ذلك العرض الرقمي الرمزي لا يلقي ترحيبًا كافيًا من قارئ قد لا يكون ملمًا إلمامًا كافيًا بمعاني مثل تلك الأرقام والرموز، وذلك يحدث على وجه الخصوص في مجالات العلوم التي تعتمد على الكلمات والتعبيرات اللفظية أكثر منها على العمق في دنيا أرقام ورموز غامضة ! ومن أجل التغلب على مثل تلك المشكلة، كان لزامًا على الباحث أن يتبع ذلك العرض الممثل للنتائج التحليلية المستخرجة على حالها، بعرض لفظي توضيحي لما يقصد بما تم عرضه من أرقام ورموز.

و - **التوضيح اللفظي لنتائج التحليل Verbal Explanation**: ويتم ذلك على خطوتين رئيسيتين أولهما تتمثل في توضيح المعاني المعنية من خلال العرض العام للدراسات المتخصصة ثم يتم الاستعانة بذلك الشرح النظري في إجراء الخطوة الثانية التي يتم من خلالها التوضيح اللفظي للنتائج التحليلية المستخرجة.

الخطوة الأولى: عرض نظري للمعاني اللفظية لبعض الدلائل الإحصائية المرتبطة لمعادلة الانحدار

يتمثل الشكل الأساسي للنموذج الخطي Linear model لمعادلة الانحدار بالكيفية الآتية:

$$Y = a + b_1x_2 + b_2x_2 + .. + b_nx_n + u$$

وذلك حيث أن:

- (Y) يمثل المتغير التابع dependent.
- $(X_1 - X_n)$ مجموعة المتغيرات المستقلة Independent أو التفسيرية explanatory التي يفترض تأثيرها المعنوي على المتغير التابع، نتيجة لما ما حدث من تغييرات في قيم أو اتجاهات المتغير التابع أو كليهما كلما حدث تغيير في قيم أو اتجاه المتغيرات المستقلة أو كليهما.
- (a) الثابت Constant أو Intercept ويمثل الرقم المطلق في المعادلة مشيرًا إلى الجزء المقطوع من محور الصادات حيث يبدأ من بعده الخط الممثل لمعادلة الانحدار المعنية. فإذا كانت قيمة الثابت صفرًا فإن ذلك يعني أن خط الانحدار سيمر بنقطة الأصل. وذلك الرقم المطلق يمكن أن يشير إلى القيمة الدنيا للمتغير التابع التي تتحقق له بدون أي تأثيرات من قبل المتغيرات المستقلة وقيمها المتغيرة.
- $(b_1 - b_n)$ معاملات المتغيرات المستقلة Coefficients وهي تمثل التغير في قيمة المتغير التابع نتيجة لحدوث تغيير قياسي واحد في قيمة المتغير المستقل، كما أنها تمثل ميل انحدار المتغير التابع على المتغير المستقل. وبناءً عليه فإن:

b_1 يمثل ما يحدث من تغيير في قيمة المتغير التابع Y نتيجة لتغيير قيمة المتغير المستقل X_1 بوحدة واحدة كما أنه يمثل ميل انحدار المتغير التابع Y على المتغير المستقل X_1 .

B_2 يمثل ما يحدث من تغيير في قيمة المتغير التابع Y نتيجة لتغيير قيمة المتغير المستقل X_2 بوحدة واحدة كما أنه يمثل ميل انحدار المتغير التابع Y على المتغير المستقل X_2 .

B_3 يمثل ما يحدث من تغيير في قيمة المتغير التابع Y نتيجة لتغيير قيمة المتغير المستقل X_3 بوحدة واحدة كما أنه يمثل ميل انحدار المتغير التابع Y على المتغير المستقل X_3 .

- (u) المتبقيات أو المتخلفات Residuals: وهي تمثل قيمة تأثير العوامل أو المتغيرات التي لم يتم إدخالها داخل النموذج أو داخل معادلة الانحدار المتعدد موضع التحليل (افتراض ثباتها لدواعي التبسيط التحليلي، أو نتيجة للجهل بكنهها أو بمدى تأثيرها الفعلي على المتغير المعني).

الخطوة الثانية: توضيح لفظي لمعاني بعض الاختبارات الإحصائية شائعة الاستخدام

- (R) هو معادل الارتباط البسيط الذي يقيس قوة العلاقة بين متغير تابع وآخر مستقل كما انه يوضح مقدار التباين في قيم المتغير الذي يرجع كنتيجة لتأثيرات المتغيرات المستقلة وقيمها عليه. مثلاً إذا كانت قيمة R هي 1.6٪ فذلك يعني أن 1.6٪ من التباين الحادث في قيم المتغير التابع (Y) ترجع لتأثيرات قيم المتغيرات المستقلة.

- (R^2) هو معامل التحديد الذي يستخدم لمعرفة القوة التفسيرية للنموذج أو للمعادلة المقدرة في حالة الانحدار الخطي البسيط (متغير مستقل واحد مع متغير تابع واحد).

- (R^{-2}) يستخدم لتفسير القوة التفسيرية لنموذج الانحدار الخطي المتعدد، لأنه يأخذ في الاعتبار (عدد) المتغيرات المستقلة.

ويتراوح قيمة R^{-2} بين 0 و 1 وكلما اقتربت قيمته من الواحد صحيح فذلك يشير إلى أن العلاقة تامة بين المتغيرات المستقلة أو المفسرة من ناحية وبين المتغير التابع من ناحية أخرى، وأنه لا توجد متغيرات أخرى تتشارك (مع المتغيرات الموجودة في المعادلة) في تأثيراتها ؛ وإن المتغيرات التفسيرية بالمعادلة تمثل المتغيرات التفسيرية 100٪ أي تمثيلاً كاملاً.

- (T) اختبار إحصائي يستخدم من أجل الحكم على معنوية معاملات الانحدار وعلى مستوى الاحتمالية المقابل له.
- (F) درجات الحرية للحكم على معنوية النموذج المقدر ككل عند مستوى معنوية معين.
- تستخدم «بيتا المعيارية» لإيجاد معادلات الانحدار المفترضة.

بينما تستخدم بيتا غير المعيارية (الحد الثابت) لتحديد معاملات المسار Path coefficients وهي أفضل في الاستخدام من بيتا المعيارية لأنها أكثر في معنوية القياس والدلالة.

ثالثا: برنامج SPSS وأهم خطوات استخدامه التصنيفية والتحليلية

(1) النشأة التاريخية للبرنامج وتطور إصداراته

يعتبر ذلك البرنامج من أكثر البرامج استخدامًا وتطبيقاً ليس فقط في المجالات الاجتماعية لكن أيضًا في المجالات العلمية الأخرى وذلك لاتساع عملياته ودوائر نتائجه ولسهولة المطلوب من إجراءات وخطوات تنفيذه. وذلك البرنامج اسمه بالكامل هو «الحزم الإحصائية للعلوم الاجتماعية Statistical Package for Social Sciences». ولقد صدر أول إصدار له في عام 1975 وكان يعمل وقته على أجهزة الحاسوب المركزية Main Frame ثم ظهرت نسخة مطورة منه هي «$SPSS^{x}$» في عام

1983، وكانت تعمل بنفس نظام البرنامج الأساسي لكنها تميزت بسهولة أكثر في الاستخدام وبسرعة أكبر في الانجاز مقارنة بالنسخة السابقة لها وفي عام 1990، صدرت أول نسخة من ذلك البرنامج صالحة التشغيل على الحاسب الشخصي PC وكان اسمها SPSS/PC 90+. وفي عام 1993، صدرت نسخ متطورة على نظام النوافذ Windows، وتوالت بعدها الإصدارات المتطورة على فترات سنوية على وجه التقريب.

(2) الخطوات الرئيسية لتشغيل البرنامج

أ - أضغط على Start ثم اختار Programs.

ب- اختار Windows for SPSS ثم اختار Windows for SPSS 11.0 فيتم فتح البرنامج.

ج- عند تشغيل البرنامج تفتح نافذة «تحرير البيانات» edit. وهي شبكة تتكون من صفوف وأعمدة لتسجيل وحفظ البيانات المزمع استخدامها في التحليل فيها . كل صف في تلك الشبكة يمثل حالة أو مفردة أو مشاهدة (مثل أفراد المجتمع محل الدراسة). وكل عمود يمثل متغير Variable.

- وهناك نافذة أخرى تستخدم لعرض المتغيرات مكتوب عليها « Variables View» وهي تخصص لتسمية المتغيرات وتحديد مستوياتها وترتيباتها.
- ويمكن التنقل بين نافذة تحرير البيانات data view، ونافذة تحرير المتغيرات، وذلك بضغط الفارة عند الاسم المختار للنافذة والمدرج أسفل الشاشة.
- وبعد إتمام إدخال البيانات وإعطاء الأوامر الملائمة لتنفيذ عمليات التحليل والاختبارات المطلوبة (التي سنشرحها بالتفصيل في السطور التالية)، سوف تظهر نتائج ذلك التحليل في ما يطلق عليها بنافذة المخرجات Output SPSS viewer.

د - تعريف المتغيرات وعرضها Variable view

- لابد من تعريف نوع المتغير هل هو رقمي Numeric أو اسمي dummy or String.
- اسم المتغير يبدأ بحرف ثم يتبع بحروف أخرى أو أرقام أو رموز مثل $ أو # بشرط ألا يتجاوز اسم المتغير ثمانية رموز.
- يحظر تمامًا أن يبدأ اسم المتغير بشرطة أو بنقطة، وألا يحتوي الاسم على فراغات أو رموز لها استخدامات خاصة مثل * أو !.
- عدم تكرار اسم متغير لكي يسمى به متغير آخر فذلك يتسبب في الخلط في المفاهيم وبالتالي في العمليات التحليلية المنفذة.
- بعض الأسماء أو تجمعات حروف معينة يمكن أن يرفضها البرنامج خاصة إن تشابهت مع معان معروفة مثل ne, get, eq, not, it, of, all, on, or, with, by, thru.

(3) مثال تطبيقي على تحليل انحدار متعدد باستخدام برنامج SPSS

إذا أردنا إجراء بحث تقييم تحليلي جاد **بهدف** «**تحديد أكثر الأنظمة الاقتصادية نجاحًا**» ولدواعي تسهيل عملية التقييم المقارن، فسوف نحصر نطاق الدراسة في أكثر النظم الاقتصادية شيوعًا في التطبيق (وهي الرأسمالية والاشتراكية) أو في الاستهداف للتطبيق المستقبلي (وهو النظام الإسلامي). وبعد التعمق في الدراسات التي تناولت كل ما يرتبط بهدف الدراسة ونطاقه، وفي التعرف الشامل على المناهج المختلفة للبحث العلمي وأدواته، تقرر اختيار توليفة مختلفة من المناهج العلمية بحيث تتلاءم مع طبيعة محور الدراسة المتمثل في «النظم الاقتصادية بين المتحقق والمأمول». وبناء عليه، فقد تقرر تطبيق المنهج التاريخي الذي يفيد في استعراض نشأة وتطور النظم المعنية، والمنهج القياسي الذي يفيد في تحديد المتغيرات التي يمكن استخدامها للتمثيل الأفضل لكل نظام من ناحية، وتلك المتغيرات التي يتوقع أن

تكون لها دلالات ارتباطية تقييمية معنوية على درجة نجاح أو فشل كل نظام بما يخدم تحقيق الهدف المرجو. ولقد تقرر استخدام نموذج الانحدار المتعدد بالاستعانة ببرنامج SPSS وذلك لتلاؤم كل مع التحليل الإحصائي المستهدف تطبيقه والذي يتناول مجموعة من التأثيرات المختلفة على كل نظام من النظم محل الدراسة. وفي تلك الحالة سوف يكون التحليل آني في ذات لحظة التحليل حيث أن المقارنة بين مجموعة مختلفة من الأشياء يتطلب بالضرورة توحيد الوحدة الزمنية للمقارنة وإلغاء تأثير عنصر الزمن الذي تتفاوت فيه درجات نجاح النظم وفقًا لأسباب مختلفة مثل الحروب وانتصارات بلاد تفرض أنظمتها وهزيمة بلاد ترضخ لأنظمة الغير حتى إن كانت ما تطبقه من أنظمة أكثر نجاحًا في الأصل.

وتتجلى **أهمية البحث** في أنه في الآونة الأخيرة شهد العالم صراعًا قويًا متناميًا بين الحضارات وبالتالي بين الأنظمة المختلفة، وبعد انهيار الاتحاد السوفيتي المهيمن على النظام الاشتراكي في نهاية الثمانينيات، تربع النظام الرأسمالي على عرش التطبيق العالمي تحت مسمى «النظام العالمي الجديد». ونتيجة لوحدانية القطبية الأمريكية المهيمنة على العالم تم فرض النظام الرأسمالي (المسمي أيضًا بنظام السوق الحر) على كل دول العالم بما فيها الدول الإسلامية والنامية والدول الاشتراكية ذاتها. وأدى ذلك إلى حدوث أزمات اقتصادية عالمية متفاقمة مما دفع الشعوب إلى إشعال الثورات العنيفة مثلما حدث في دول أمريكا اللاتينية والصين ثم حاليًا في منطقة الدول العربية، وقد تم ذلك ليس فقط لأسباب اقتصادية، ولكن أيضًا لدواعي سياسية ونتيجة للتصارع على استعادة كل نظام لهيمنته السابقة. وفي خضم تلك الأحداث، بزغ العملاق الإسلامي في محاولة لاستعادة مجد الدولة الإسلامية القوية في فترة ما من الزمن الماضي.

وقد أدت الثورات والأزمات العالمية وأحداث ما قبل ثورات الربيع العربي إلى إعادة التفكير في إنشاء نظام اقتصادي سياسي مستحدث نعت بالطرف الثالث حيث استهدف له أن يكون خليطًا من إيجابيات النظامين المهيمنين على العالم وهما النظام الرأسمالي والنظام الاشتراكي. أما ثورات الربيع العربي فأصبح يظللها مدًا

إسلاميًا لا يمكن تجاهله، وأصبح ذلك العملاق المتوالد من جديد يستهدف تطبيق نظام إسلامي شامل.

من تلك التوجهات الأخيرة تبرز أهمية التحليل الحالي الذي يهدف إلى تقديم صورة تقييمية موضوعية للنظم الثلاث: الرأسمالي، الاشتراكي، والإسلامي من منظور مقارن حيادي، لعل الصورة التحليلية المقدمة تسهم في المساعدة على تحديد أفضل لطبيعة النظام الشامل الواجب إرساء لبناته وتنمية هيكله الفاعل في تلك الفترة الانتقالية الهامة من تاريخ العالم المعاصر.

وبناء عليه، فسوف يتم النظر إلى النظام الاقتصادي موضع المقارنة باعتباره متغيرًا تابعًا حيث يتم قياسه بمؤشر يمثل تمييزا مشتركاً بين الأنظمة الثلاثة وهو حجم مساهمة كل من القطاع الخاص والقطاع العام في ظل كل من الأنظمة الثلاثة، حيث يتم اختيار ثلاث دول إحداهم يتبع النظام الرأسمالي (وبالتالي يكون القطاع الخاص هو المهيمن على الأنشطة الاستثمارية في الدولة)، وثانيها دولة اشتراكية (وبالتالي يكون القطاع العام هو المهيمن على الأنشطة الاستثمارية في الدولة). أما الدولة الثالثة فسوف تكون من الدول التي أثبتت نجاحًا في تطبيق النظام الاقتصادي الإسلامي ولو بشكل جزئي ولتكن ماليزيا (حيث يتوقع أن يتواجد فيها كل من القطاعين العام والقطاع الخاص معًا نتيجة لأن الإسلام دائمًا وسطيًا وغير متحيز لقطاع على حساب آخر.

أما المتغيرات التقييمية التي نود إدراجها في الطرف الأيسر من معادلة الانحدار باعتبارها متغيرات مستقلة وتفسيرية للنظام المذكور فسوف تتمثل في تلك العوامل التي يمكن اعتبارها مقاييس ومؤشرات على نجاح أي نظام، ولدواعي التبسيط نختار منها متغير «الديمومة» (وهو مقياس زمني يتم قياسه بعدد سنوات استمرارية النظام) ومتغير «الشمولية» (وهو مؤشر مكاني يتم قياسه بعدد المناطق المطبقة له) ومتغير الحالة الاقتصادية للأفراد (ويمكن قياسه بمتوسطات دخول الأفراد العائلين لأنفسهم أو للغير).

وقبل الاستطراد، يلزم التنويه إلى أننا نعرض مثال تصوري قد يعيبه عدم الدقة في اختيار المقاييس الصحيحة والمتغيرات الممثلة لكنه يعرض الآن فقط بهدف شرح طريقة تطبيق أسلوب الانحدار المتعدد باستخدام برنامج الـ SPSS.

كما أن هناك ملاحظة أخرى حول ما سوف يعرض من خطوات. وهي أنها تمثل صورة عامة يمكن أن تختلف في بعض جزئياتها نتيجة إما لاختلاف أجهزة الحاسوب واختلاف الإمكانيات الفعلية المتاحة في كل جهاز، بحيث يمكن أن تكون غير صالحة إجمالاً لتطبيق البرنامج التحليلي المذكور أو أنها تتطلب ويندوز معين. كما قد يحدث الاختلاف نتيجة لنسخة البرنامج حيث أوضحنا سلفًا أنها تتطور بشكل متلاحق منذ عام 1975. وقد تكون هناك أسباب أخرى تتطلب الاستعانة بخبير برمجيات أو متخصص إحصائي أو الاثنين أو خبير ثالث له خبرة أوسع في ذلك المجال التحليلي خاصة أنه في بعض الأحوال يتم رفض قبول رموز للأفعال إذا كانت تخرج عن نطاق عدد الأحرف أو الأرقام المسموح بها.

وبوجه عام، فإنه برغم التقدم التقني والبحثي الملموس في هذا الصدد، فمازالت العمليات التنفيذية يشوبها كثير من الصعوبات ويعترضها كثير من الوقفات التي تتطلب مساعدة خبراء أو زملاء يقومون بذات العمل البحثي التحليلي من أجل تحقيق عمل متكامل ودعم الخبرات المتعددة ولتعتبر الخطوات التالية هي دليل عام مرشد.

الخطوة الأولى: التحديد القياسي لكل من المتغير التابع والمتغيرات المستقلة بشكلها النهائي التي سوف يتم تحصيل بيانات بعينها لقياسها وفقًا لذلك التحديد

وفي مثالنا الحالي ذكرنا أن المتغير التابع يتمثل في النظام الاقتصادي الذي نبحث في درجة فعاليته مقارنة بالنظم الأخرى الشائعة. فإذا افترضنا أن النظم الاقتصادية محل المقارنة هي النظام الرأسمالي، والنظام الاشتراكي والنظام الإسلامي، وأن أهم متغير أو وحدة قياسية يمكن أن تميز كل نظام عن الآخر بحيث تصبح مؤشرًا تمييزيًا

أفضل هي مدى مساهمة القطاع الخاص أو القطاع العام، فذلك في حد ذاته لا يصلح لبداية القياس المستهدف، وبالتالي علينا اختيار مؤشر قياسي يصلح توحيده وجمع بيانات ملموسة حوله. ولقد وجد أن ذلك سوف يتمثل في «حجم الاستثمارات» للقطاع الخاص مقدرًا بالعملة الوطنية (وليكن بالجنيه المصري). ذلك حيث أنه إذا زاد ذلك الحجم عن 60٪ يدل ذلك على أن النظام رأسمالي لأنه يقوم على أغلبية مساهمة للقطاع الخاص، وإذا انخفض الحجم عن 20٪ يدل ذلك على أن النظام اشتراكي لأن القطاع العام هو القائم بغالبية الاستثمارات المذكورة، ولقد تم اختيار النسبتين بشكل تحكمي افتراضيًا بأنه النظام الاشتراكي لا يقبل أبدًا أن تخصص للقطاع الخاص أغلبية الاستثمارات التي يمكن أن تمثل أكثر من نصف إجمالي حجم الاستثمارات بمعدل واضح، وعادة إذا سمح للقطاع الخاص بالقيام باستثمارات ما فيكون ذلك بمعدل ضئيل نتصور أن أقصاه 20٪ من إجمالي الاستثمارات المحلية. أما إذا وجد أن حجم الاستثمارات يتراوح بين ما يربو على 20٪ من إجمالي الاستثمارات المحلية ولكنه يقل عن 60٪ من ذلك الإجمالي فإن ذلك يمكن أن يعطينا مؤشرًا تصوريًا منطقيًا على أن النظام الذي يتحقق فيه ذلك المعدل الاستثماري هو إما اقتصاد مختلط حيث يساهم فيه كل من القطاعين الخاص والعام، أو أنه نظامًا إسلاميًا لأن من طبيعة ذلك النظام الأخير أنه نظامًا (وسطًا) يقبل كل ما عداه من نظم بشكل محايد بشرط التوافق مع التعاليم الأساسية.

والواقع أن التصور الذي ذكر توًا يمكن أن يصلح في حد ذاته ليكون فرضية تتطلب من فرق متعاقبة من الباحثين التحقق من صحته أو من عدمه، مع استنباط معايير قياسية تدعم تلك المعدلات الافتراضية التي وضعناها بشكل تحكمي استقراءً للشواهد واستنباطًا من الصورة العامة للخصائص التي يتميز بها كل من النظم الاقتصادية الثلاث المعنية.

وبافتراض أن ما ذكرناه تم قبوله علميًا، فننتهي إلى اعتبار المتغير التابع ممثلاً بنسبة مساهمة القطاع الخاص في إجمالي الاستثمارات المحلية، حيث يتطلب ذلك

حصر كل الاستثمارات المحلية التي يساهم فيها القطاع الخاص بشكل إجمالي أو جزئي؛ مباشر أو غير مباشر، مع تحصيل قيم تلك الاستثمارات وتقويمها جميعها بعملة موحدة يفضل أن تكون العملة الوطنية لتجانس طبيعتها مع طبيعة الاستثمارات المعنية وهي (المحلية) فقط. وتعد قائمة طولية بتلك القيم.

وعلى ذات النهج يتم تحديد المؤشرات القياسية لكل من المتغيرات المستقلة المقرر استخدامها بنموذج الانحدار مع تحديد القيم القياسية التي تمثل البيانات المستهدف تحصيلها وإدراجها في المكان المحدد في محرر البيانات باستخدام الحاسوب الملائم. ويتم ذلك - تصورًا- على النحو التالي:

✍ المتغير المستقل الأول هو الديمومة Contumacy:

ويمكن أن يدخل في التحليل بإحدى الطريقتين:

أ - أن يقاس كمتغير رقمي Numeric بعدد سنوات استمرارية تطبيق النظام بدون انقطاع خلال فترة زمنية معينة لتكن العشرين عامًا الماضية. حيث يتم هنا حساب عدد سنوات استمرارية نشاط كل استثمار قام به القطاع الخاص بدون انقطاع (باعتباره الممثل الافتراضي للنظام) وبناء عليه، يتم تحصيل البيانات الممثلة بذلك المؤشر القياسي من داخل كل مؤسسة من مؤسسات القطاع الخاص ثم يتم إدراج تحصلاتها النهائية في صف طولي مقابل للبيانات المرتبطة بكل فئة استثمارية قام بها القطاع الخاص على حده.

ب- أن يقاس ذلك المتغير بصفته متغيرًا صوريًا Dummy، حيث يعطي رقم (1) لكل فئة استثمارية قام بها القطاع الخاص وتحققت لها صفة الديمومة عبر الفترة المحددة، ورقم (0) أو صفر ما عدا ذلك.

✍ المتغير المستقل الثاني هو الشمولية Comprehension:

والذي يدل على عدد المناطق التي يغطيها الاستثمار الخاص، ومثل الحال في قياس المتغير السابق يمكن القيام بنفس الشيء في المتغير الحالي:

أ - فيمكن أن يقاس متغير الشمولية بصفته متغيرًا رقميًا. ويتم ذلك بتحصيل بيانات حول عدد المناطق الجغرافية (المحلية والخارجية) التي تعطيها كل فئة استثمارية قام بها القطاع الخاص وتدرج في صف طولي مقابل لسابقيه؛

ب- ويمكن أن يعتبر ذلك متغيرًا صوريًا بإعطاء رقم (1) للاستثمارات الخاصة التي لوحظ ارتفاع معدلات تغطيتها لعدد كبير من المناطق (المحلية/ الخارجية)، وإعطاء رقم (0) أو صفر لما عدا ذلك.

لكن في هذه الحالة، لابد أولاً من تحديد الفئة العددية التي تقاس بها صفة الشمولية من عدمها كأن يقال مثلاً أن الشمولية تتحقق إن ارتفعت نسبة تغطية الاستثمارات الخاصة عن 75٪ من المناطق المحلية لا الأجنبية، (على مستوى المحافظات أو على مستوى المحليات أو ... أو ...) وتنتفي صفة الشمولية عن الاستثمارات الخاصة التي تتدنى معدلات تغطيتها عن 75٪ من المناطق المستهدف توزيع الخدمات عليها أو المتوقع تسويق منافعها من خلالها (أو ... أو ... أو ..).

✍ **المتغير المستقل الثالث الافتراضي هو الحالة الاقتصادية للأفراد العائلين لأنفسهم أو لآخرين**

ويتم تبرير اختيار تلك الفئة لأنها التي يفترض امتلاكها للقوى الشرائية الحقيقية المطلوبة من أجل سد الحاجات الاتفاقية والمعيشية لأفراد المجتمع ككل نتيجة لصفة (الإعالة) التي يتفردون بل ويتكلفون بها.

وذلك المتغير يمكن قياسه بالعديد والعديد من المؤشرات المستقلة أو التكاملية مثل متوسط المنفق على الطعام، أو قيمة الحيازات الثابتة، أو الحالة المهنية والتعليمية ... الخ. لكن مازال متوسط دخل الفرد يمثل المعيار الرئيسي المفضل لقياس الحالة الاقتصادية فهو يمثل القاسم المشترك الأعظم في تمثيل درجات النمو والتنمية الاقتصادية والاجتماعية على جميع المستويات.

وبناء عليه، فإن المتغير الثالث يتحدد بمؤشر «متوسط دخل الفرد» لفئة الأفراد

العائلين لأنفسهم أو لذويهم أو لآخرين، مقومًا بالعملة المحلية (التي تمثل القوة الشرائية المحلية). وسوف يتم تحصيل متوسط داخل الفرد الشهري أو السنوي، بحيث يعتبر في حد ذاته مؤشرًا على نجاح وفعالية النظام، فكلما ارتفعت متوسطات دخول الأفراد والعائلين لأنفسهم أو لذويهم أو لآخرين... وكلما تقاربت تلك المتوسطات كلما دلّ ذلك على نجاح النظام المطبق وعلى فعاليته في النهوض بالمجتمع ككل والعكس بالعكس بالتأكيد سيكون صحيحًا. وأيا ما كان الحال، فسوف يتم تحصيل البيانات الممثلة لذلك المتغير وتدرج في عمود رابع مقابل لما سبقه.

الخطوة الثانية: بعد وضع البيانات بقياساتها المحددة بدقة في الجدول اليدوي، يتم إدخالها في محرر البيانات Data File على ذات الهيئة الموضحة بالجدول التالي:

جدول (3) البيانات الرقمية القياسية للمتغير التابع والمتغيرات المستقلة في معادلة الانحدار

Input - SPSS Data Editor							
File Edit View Data Transform Analysis Graphs utilities Window help							
	Y	X_1	X_2	X_3	Var	Var	Var
1	100,000	1	15	1000			
2	559,00	1	8	3000			
3	1,488,50	∅	7	10,000			
4	7,999,000	1	10	17,750			
5	10,400,000	1	25	5,120			
.							
.							
.							

ملحوظات: (1) الأرقام كلها افتراضية.

(2) متوسطات الدخول للأفراد العائلين من العاملين بالاستثمارات الخاصة المقابلة لها بالجدول الموضح.

الخطوة الثالثة: نذهب إلى قائمة المدخلات **Inputs** (داخل المحرر الفرعي للبيانات)، حيث تتم تسمية المتغيرات المدرجة بالجدول السابق وتحديد المعنى اللفظي القياسي لكل منها حتى يتم تعريفها في بيان مخرجات النتائج النهائية للتحليل المعني، وذلك كما يتضح في الجدول التالي:

جدول (4) تسمية المتغيرات الداخلة في النموذج التحليلي وتحديد طبيعتها (رقمية أو صورية)

Input - SPSS Data Editor				
File Edit View Data Transform Analysis Graphs utilities….				
Serial	Name	Type	Width	Decimals
1	Y	Numeric	8	2
2	X_1	Dummy	8	2
3	X_2	Numeric	8	2
4	X_3	Numeric	8	2

ملحوظة: تم تحديد المتغير الأول (الديمومة) ليعبر عنه كمتغير صوري على سبيل الافتراض لا التمثيل الواقعي.

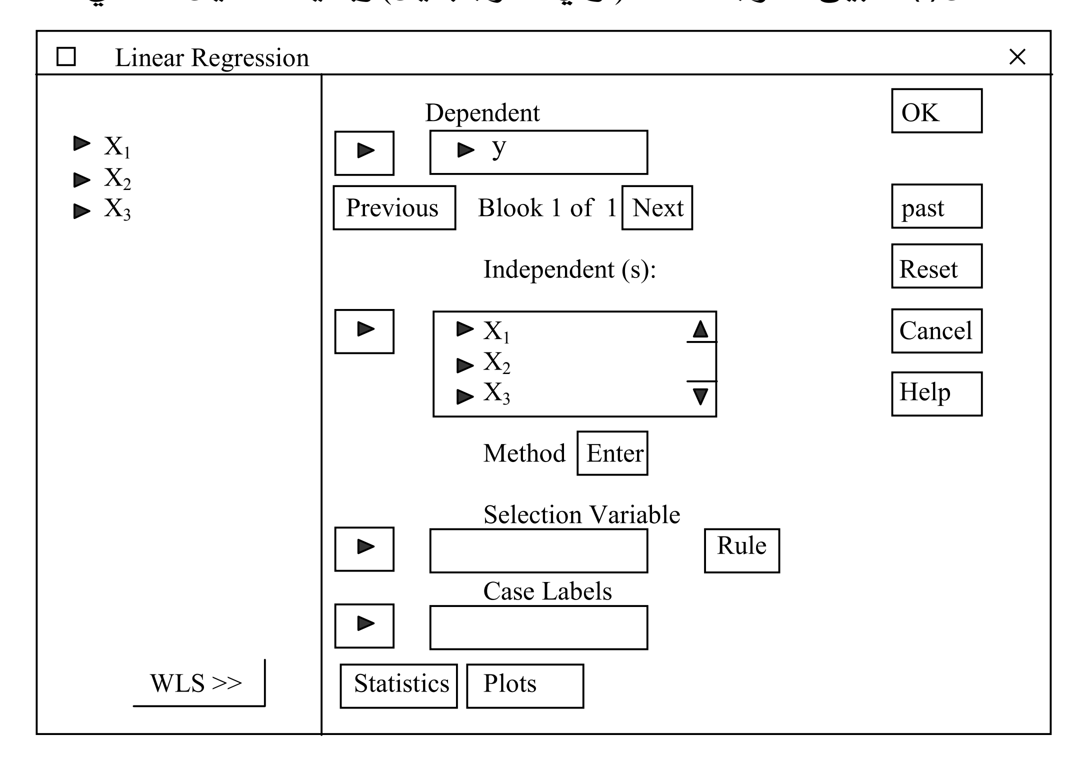
الخطوة الرابعة: نبدأ بأولى الخطوات الدخول في العملية التحليلية المستهدفة حيث نذهب إلى قائمة «تحليل» **Analysis** ونختار منها أمر **Command** «الانحدار» **Regression** ومن القائمة الفرعية المصاحبة نختار أمر «خطي» **Linear** وذلك على النحو التالي:

جدول (5) تحديد أسلوب التحليل المستهدف (الانحدار الخطي في مثالنا الجاري)

Input - SPSS Data Editor					
File Edit View Data Transform			Analysis Graphs utilities	Window help	
			Reports ▶		
			Descriptive Statistics ▶		
			Compare Means ▶		
			General Linear Model ▶		
	Y	X_1	Correlate ▶	Var	Var
1	100,000	1	Regressicvl ▶	Linear	
2	559,00	1	Classify ▶	Curve Estimation	
3	1,488,50	∅	Data Reduction ▶		
4	7,999,000	1	Scale ▶		
5	10,400,000	1	Nonparametric Tests ▶		
			Multiple Response ▶		

الخطوة الخامسة: في نفس نافذة تحليل الانحدار **Analysis** يتم الفصل بين المتغير التابع من ناحية والمتغيرات المستقلة من ناحية أخرى، حيث يتم تحديد المتغير التابع (Y) وهو ما يتمثل بحجم الاستثمارات المحلية التي يقوم بها القطاع الخاص، ثم يتم نقل قيمه إلى خانة المتغير التابع. ثم يتم تحديد المتغيرات المستقلة وتنقل إلى خانة المتغيرات المستقلة ثم تثبت العملية في وضعها التصنيفي النهائي بنقر مفتاح (OK). وفي نفس الخطوة يتم اختيار وتحديد أسلوب التحليل التطبيقي المزمع تطبيقه وهو عادة واحدًا من خمسة أساليب **Forward, Backward, Remove, Stepwise, Enter**. وسوف يتم توضيحها بعد عرض المعروض على نافذة تحليل الانحدار بعد استكمال الخطوة الحالية.

شكل (3) تطبيق أسلوب «**Enter**» (أو أي أسلوب بديل) وبداية التحليل العملي



الخطوة السادسة: طباعة المخرجات التي ستظهر في عدة أقسام كما يتضح من الأشكال الجدولية المتتالية الآتية:

جدول (6) نتائج تحليل الا نحدار الخطي المتعدد التي تظهر على شاشة الحاسوب

Regression (أ6)

Model	Variables Enter	Variable Removed	Method
1	X_1 X_2 X_3		Enter

Std. Standard deviation

Model Summary (ب6)

Model	R	R Square	Adjusted R Squire	Std Error of the Estimate
1	975	951[0]	938	452761

NOVA (ج 6)

Model	Sum of Squares	df(1)	Mean Square	F	Sig(2)
1 Regression	4374508	3	1458169	71133	000[a]
Residual	225492	11	20499		
Total	4600000	14			

(1) Degrees of Freedom

(2) Significance

a - predictors (constants) X_1, X_2 , X_3

b - Dependent variable y

Coefficients (د 6)

Model	Unstandardized coefficients		Standardized coefficients	T	Sig
	B	Std error	Beta		
1 (Constant)	79106	19782		3.999	002
X_1	-4928	1611	-563	-3.059	011
X_2	1.590E.02	007	+392	2.146	055
X_3	175	637	-510	1	

الخطوة السابعة: شرح وتفسير لفظي لمخرجات التحليل

قبل شرح النتائج الرقمية والرمزية بالعرض اللفظي الصحيح، يجب أولاً إعداد قائمة هامشية لتوضيح المفاهيم محل الشرح والتوضيح. ولقد عرضنا بعضها في جزء سابق يرتبط بمعاني مفردات دالة الانحدار المتعدد (الخطية)، ومعاني بعض رموز الاختبارات والدلالات المعنوية الإحصائية ويتبقى لنا أن نعرض بإيجاز شرحًا للأساليب التحليلية التي تستخدم في تطبيق تحليل الانحدار الخطي المتعدد وفقًا لبرنامج SPSS وذلك على النحو التالي:

✍ **أساليب تطبيق الانحدار وتحليله Methods of Regression Analysis**

(1) Enter:

ويقوم ذلك الأسلوب على إدخال كل المتغيرات المستقلة مرة واحدة مع المتغير التابع.

(2) Stepwise:

إدخال المتغيرات المستقلة بشكل متدرج. ففي خطوة أولى يتم إدخال المتغير التابع مع متغير واحد مستقل يفترض أنه الأكثر قوة في ارتباطه بالمتغير التابع حيث يستدل على ذلك بإجراء عمليات سابقة لتحليل الارتباط البسيط والتعرف على درجة ارتباط كل متغير مستقل على المتغير التابع على حده.

وفي خطوة ثانية يتم إدخال متغير ثانٍ ذي قوة ارتباط تالية في درجتها للمتغير المستقل الذي سبق إدخاله وتستمر عمليةٍ إضافة بقية المتغيرات المستقلة واحدًا تلو الآخر وفقًا لقوة درجة الارتباط بالمتغير التابع. ويمكن من خلال تتبع اختبارات المعنويات (مثل R^{-2}) التعرف على أكثر مجموعة من المتغيرات المستقلة قوة في تأثيراتها مجتمعة على المتغير التابع واستبعاد المتغير الأضعف تأثيرًا لأن ذلك الضعف يشير إلى عدم صلاحيته للإدخال في النموذج التحليلي.

وكذلك، فإنه بعد إكمال إدخال جميع المتغيرات المستقل، يمكن تكرار العملية

في اتجاه عكسي وذلك باستبعاد متغيرًا مستقلاً تلو الآخر وذلك وفقًا للأضعف تأثيرًا فما يفوقه قوة لكنه مازالت أضعف من غيره في تأثيره على المتغير التابع.

(3) Remove:

يتم إدخال كل المتغيرات المستقلة مرة واحدة ثم يتم حذف المتغيرات المستقلة ذات الارتباط الصفري أو الضعيف دفعة واحدة.

(4) Backward:

يتم إدخال كل المتغيرات المستقلة مرة واحدة ثم يتم حذف المتغير المستقل الأقل في درجة ارتباطه الجزئي (البسيط) مع المتغير التابع، ثم يحذف ما يليه في ضعف درجة ارتباطه وهكذا حتى يبقى في المعادلة فقط المتغيرات المستقلة ذات المعنويات الإحصائية المرتفعة.

(5) forward:

عكس الأسلوب السابق؛ يدخل المتغير التابع مع المتغير المستقل ذي الارتباط الأعلى (البسيط الجزئي) معه ثم تكرر العملية مع التالي في درجة ارتباطه وهكذا حتى يتم الوصول إلى الحد الذي لا تحدث معه أية زيادة في معامل الارتباط.

هذا، ويعتبر أسلوب Enter و Stewise هما الأكثر شيوعًا في الاستخدام.

العرض اللفظي التوضيحي للمعروض في بنود وقوائم التحليل المطبوعة

(1) توضيح الأسلوب المستخدم في التحليل على النحو التالي: تم استخدام الأسلوب التحليلي Enter حيث أدخلت جميع المتغيرات المستقلة معًا في آنٍ واحد في معادلة الانحدار الخطي المتعدد وذلك كما يتضح في جدول (6 أ).

(2) من المخرجات الموضحة في جدول (6 ب)، يتضح الآتي

أ - معامل الارتباط البسيط R=0.975، وهو ما يدل على وجود ارتباط قوي بين المتغير التابع والمتغيرات المستقلة المدرجة بالمعادلة أو بكلمات أخرى،

فذلك يشير إلى وجود تأثير معنوي قوي من المتغيرات المستقلة X_3, X_2, X_1 على التباين الذي يحدث في قيم المتغير التابع Y (المتمثل هنا في حجم استثمارات القطاع الخاص. وذلك لأن قيمة المعامل تقترب من 100٪ وهي درجة الارتباط القصوى أو التامة بين المتغيرات.

ب- معامل التحديد R^{-2}=0.951 ، أيضًا تقترب تلك القيمة من الرقم النموذجي 100٪، الذي يرتبط في تلك الحالة بدرجة القوة التفسيرية للمعادلة المقدرة وذلك في حالة الانحدار الخطي البسيط (متغير مستقل واحد مع المتغير التابع).

ج- معامل التحديد المصحح R^{-2}=0.938 ، وهو كما أوضحنا من قبل يستخدم للدلالة على القوة التفسيرية لنموذج الانحدار الخطي المتعدد (وليس البسيط) حيث يوضح قوة النموذج بكل المتغيرات المستقلة الداخلة في إطاره ككل. والرقم الإحصائي الناتج كقيمة لمعامل التحديد المصحح (R^{-2}) تقدر في مثالنا الحالي بقيمة مرتفعة تقارب 100٪ لأنها تبلغ 0.938 ومن ثم فهي تدل على أن مجموعة المتغيرات المستقلة تمثل النموذج تمثيلاً قويًا ولأنها تفسر ما يحدث من تباين في قيم المتغير التابع (وهو هنا حجم استثمارات القطاع الخاص) بشكل قوي بمعنى أن ما يحدث من تباين في قيم المتغير التابع يرجع إلى أكثر من 90٪ إلى تغييرات في قيم المتغيرات المستقلة الداخلة في النموذج.

د - أما بقية التأثيرات على تباين قيم المتغير التابع (والتي تبلغ هنا 0.06 وتمثل بما يطلق عليه البواقي أو المتبقيات Residuals)، فيرجع حدوثها إلى عوامل أخرى أو متغيرات أخرى لم تؤخذ في الاعتبار في المعادلة المحسوبة أو المقدرة.

(3) بالانتقال إلى البيانات الموضحة في جدول (6 ج)، نجد أن ذلك القسم من

المخرجات التي تعرض بهيئتها هذه على الشاشة وتمت طباعتها على ذلك الحال، يتضمن قيم التباين الذي يساعد الباحث التحليلي على أن يتعرف باستقراء ما فيه من مخرجات على القوة التفسيرية للنموذج ككل وذلك من خلال اختبار F أو F test.

وبالاستعانة بجدول تحليل التباين لمعنوية F، لوحظ المعنوية العالية لاختبار F ($P<0.0001$) مما يؤكد القوة التفسيرية العالية لنموذج الانحدار الخطي المتعدد من الناحية الإحصائية.

(4) أما الجدول رقم (6 د) فيمكن من خلاله التعرف على قيمة الثابت (a) الموضحة بمعادلة الانحدار الخطي المتعدد المقدرة، وكذلك التعرف على معادلات الانحدار الخاصة بكل المتغيرات المستقلة (B_S) والتعرف على درجات تأثير كل من المتغيرات المستقلة الداخلة في النموذج على القيم المختلفة للمتغير التابع، حيث يفضل إعادة ترتيب تلك النتائج الإحصائية أو ذلك النوع من المخرجات على النحو الموضح في الجدول التالي:

جدول (7) معاملات الانحدار للمتغيرات المستقلة بمعادلة الانحدار

المتغير التابع	المتغيرات المستقلة			
Y	الحد الثابت غير المعياري B	X_3	X_2	X_3
قيمة المعامل	79.106	- 4.928	0.515	0.175
قيم اختبار T	3.999	- 3.059	2.146	0.275
المعنويـــة	0.002	0.01	0.055	0.789

ويوضح اختبار t (عند مستوى معنوية $P\leq0.05$)، أن المتغير المستقل X_1 كان معنويًا من الناحية الإحصائية، بينما كان المتغير المستقل الثاني X_2 قريبًا من وضع المعنوية (عند مستوى معنوية $P\leq0.05$) ولكن عند ذات المعنوية $P\leq0.05$ وجد أن

المتغير المستقل الثالث X_3 ليس له تأثير معنوي على المتغير التابع في نموذج الانحدار الخطي المتعدد المحسوب.

ونظرًا للأهمية النسبية لذلك الأسلوب التحليلي الأكثر شيوعًا في الاستخدام، فسوف يتضمن الكتاب الحالي مثالين آخرين، تم تطبيق أحدهما بالكامل بواسطة المؤلفة في أحد البحوث المنجزة لها، وتم إعداد الآخر بواسطة الباحثة «سلمى فاروق» معيدة الاقتصاد بكلية التجارة بجامعة الأزهر أثناء دراستها العملية في فترة إعدادها لنيل درجة الماجستير.

وتجنبًا للملل وللتيه نتيجة الإطالة في مناقشة ذات الموضوع، فسوف يعرض المثالين الإضافيين في ملحق الكتاب.

الفصل الثاني

المنظور والعرض البياني Graphics

3/ 2/ 1 مقدمة:

تعتبر الأشكال البيانية Graphics and Figures أداة لا غنى عنها indispensable ولابد من استخدامها في أي دراسة علمية لكي توضح البيانات والمعلومات والنتائج بعرض مرئي يقرب الصور الكمية أو النتائج المتداخلة أو المختلطة إلى الأذهان. وهي تعتبر بالنسبة لكثير من المتخصصين (مثل الباحثين في علم الاقتصاد) بمثابة الشاكوش بالنسبة للنجار as a hammer to a carpenter.

وتعتبر الجداول tables هي التوأم الملاصق لتلك الأشكال البيانية التي عادة ما تتواجد معها أينما حلَّت، لكنها تسبقها في الظهور والعرض. ذلك ما سوف يتم توضيحه وشرحه عبر الصفحات التالية.

($E_2 = 0$) حينئذٍ ستكون معادلة لامبيدا ببساطة:

$$\lambda = \frac{E_1}{E_1} = 1$$

ب- إن قيمة لامبيدا غالباً تساوي 0 عندما تظهر البيانات أنه لا يوجد تطابق. فإذا لم يكن هناك تطابق في البيانات، فإن النتيجة المشاهدة سوف تطابق تماماً "نموذج لا تطابق"، وباستطاعتنا التكهن باستخدام "نموذج لا تطابق" الذي يولد لنا أخطاء ($E_2 = 0$). وهذه الأخطاء سوف تولد لنا قيمة 0 للامبيدا.

ج- في بعض الأحيان ستكون لامبيدا مساوية لـِ 0 عندما تُظهر البيانات بعض التطابق. مع أن لامبيدا دائماً سوف تكون مساوية لـِ 0 عندما لا يوجد تطابق. والعكس ليس دائماً بالضرورة صحيحاً، ففي بعض الأحيان عندما تكون لامبيدا مساوية لـِ 0 فإنه قد يكون حقاً هناك تطابق. وهذا يمثل القصور الأساسي في استخدام لامبيدا والذي سنشير إليه في موضعه في نهاية هذا الفصل.

د- تعتبر لامبيدا مقياساً للتطابق اللامتماثل Asymmetric. وهذا يعني أن قيمة لامبيدا سوف تكون مختلفة اعتماداً على الطريقة التي تم بها التفكير حول المتغيرين أيهما متغير مستقل وأيهما متغير تابع. بمعنى آخر، إذا ما حاولنا في المثال السابق أن نتكهن بمستويات الدخول استناداً على أين يعيش الأفراد مفضلين ذلك على الطريقة الأخرى، فإن قيمة لامبيدا ستتغير. وعليه عند استخدامنا للامبيدا ينبغي علينا أن نكون واضحين حول طبيعة العلاقة التي نعتقد أنها تربط هذين المتغيرين معاً. إن هذه الطريقة تجعل من لامبيدا طريقة مفيدة عندما يكون لدينا أسباب قوية للاعتقاد أنه توجد طريقة واحدة للعلاقة بين المتغيرين تسير في اتجاه محدد.

ولتوضيح كل ذلك دعنا نقلب المتغيرات في المثال السابق وأن نتعامل مع مكان الإقامة كمتغير مستقل، ومستوى الدخل كمتغير تابع. إن هذا الأمر يتطلب منا بناء جدول توافق حيث نضع متغير مكان الإقامة عبر الأعمدة والدخل وفقاً للصفوف (جدول 7 – 7).

جدول (7ـ7) مكان الإقامة وفقا لمستوى الدخل

مكان الإقامة	الدخل		المجموع
	منخفض	عال	
حضري	167	205	372
ريفي	230	118	348
المجموع	**397**	**323**	**720**

وبدون سابق أية معرفة لتوزيع المتغير المستقل (مكان الإقامة) فإنه باستطاعتنا التكهن بأن كل الحالات التي تقع في المجموعة ذات المستوى الاقتصادي المنخفض، لذلك فإن سوء التصنيف لـِ 348 شخصاً الذين يكسبون في الحقيقة دخلاً اقتصادياً عالياً:

$$E_1 = 348$$

وإذا ما أردنا أن نتكهن بمستويات الدخول لدى الناس، إلا أننا في هذه المرة لدينا معلومات فيما إذا كان هؤلاء يعيشون في مناطق ريفية أو حضرية، عندئذ يكون التكهن بأن كل الناس من المناطق الحضرية لديهم دخول عالية، ومن هنا نكون قد وقعنا في 118 خطأ. عليه فإن المعدل الكلي للخطأ عندما استخدمنا نموذج التطابق التام هو:

$$E_2 = 167 + 118 = 285$$

وبالتالي تصبح قيمة لامبيدا باستخدام نموذج العلاقة:

$$\lambda = \frac{E_1 - E_2}{E_1} = \frac{348 - 285}{348}$$

$$= 0.18$$

بمعنى آخر فإن التطابق يعتبر - إلى حد ما - ضعيفاً، إلا أن هذا التطابق قد يمدنا بالإحساس بأننا قادرون على أن نتوقع مستوى دخل شخص ما استناداً إلى المكان الذي يعيش فيه. فالتطابق بين هذين المتغيرين يعتبر إلى حد ما أقوى عندما نظرنا إلى مستوى الدخل كمتغير تابع بدلاً من مكان الإقامة.

إجراءات توليد لامبيدا Lambda باستخدام SPSS:

1- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Analyze ⟸ descriptive statistics ⟸ Crosstabs

2- انقر على المتغير في القائمة التي سوف تشكل صفوف الجدول، والتي تكون في هذه الحالة Place of Residence..

3- انقر على ◄ التي تشير إلى المساحة المعنونة Row(S)، تقوم بلصق Place of Residence في القائمة المحددة للصفوف Row(S) .

4- انقر على Income Level.

5- انقر على ◄ التي تشير إلى المساحة المعنونة Column(s). تقوم بلصق Income Level في القائمة المحددة للأعمدة Column(s).

6- انقر على زر Statistics، يقوم بإعطائنا صندوق Statistics: Crosstabs وفي أعلى زاوية من الشمال يمكننا رؤية المساحة المعنونة Nominal Data هذه مقاييس التطابق المتوفرة عندما يكون على الأقل أحد المتغيرات مقاساً على المستوى الاسمي. في هذا المثال Place of residence قد تم قياسه على المستوى الاسمي.

7- اختر Lambda بالنقر على الصندوق الموجود أمامها وضع علامة T في الصندوق لتبين أنه قد تم اختيار لامبيدا.

8- انقر على Continue.

9- انقر على OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

Directional Measures

	Value	Asymp std. Error[a]	Aprrox. T[b]	Aprrox. Sig
Nominal by Lambda Symmetric	.151	.047	3.061	.002
Place of residence Dependent	.118	.056	1.976	.048
Income Level Dependent	.181	.052	3.184	.001
Goodman Place of residence Dependent	.045	.015		.000 c
And Kraskal-tau Income Level Dependent	.045	.015		.000 c

a. Not assuming the null hypothesis.
b. using the asymptotic standard error assuming the null hypothesis.
c. Based on chi-sguare approximation

في هذه المخرجات تم استخدام عملية التماثل عندما لا يكون لدينا أي سبب لأن نشك في أن واحداً من هذين المتغيرين يكون معتمداً على الآخر، بقدر ما يكون هذان المتغيران - تبادلياً - معتمدَيْنِ على بعضها البعض، ففي هذا المثال قيمة التماثل تساوي 0.151 تقع في مكان ما بين القيم اللاامتماثلة 0.118 و 0.181. إن عملية استخدام اللاتماثل تكون لديها قيمتان احتماليتان استناداً على أي من المتغيرين يعتقد أنه متغير تابع.

كذلك تحتوي هذه المخرجات على قيمة متعلقة بمقياس اسمي. ولقياس التطابق ألا وهو جودمان وكرشكال تاو Goodman and Kruskal tau والذي يحتوي على قيمة صغيرة للتطابق إذا ما قورن بـِ لامبيدا Lambda في الأعمدة الأخرى من هذا المربع تحتوي على معلومات ليست ذات علاقة بهذا الفصل، وإنما ستكون هذه المعلومات ذات أهمية في مواضع أخرى من هذا الكتاب خاصة عند التعامل مع الاستدلال من العينة على المجتمع [4].

حدود استخدام لامبيدا [5]:

على الرغم من الطريقة السهلة لحساب لامبيدا إلا أن هناك مشكلة عادة ما تقابلنا عند استخدام لامبيدا. إن هذه المشاكل التي تناولناها عند الحديث عن خصائص لامبيدا يمكن أن يكون لديها قيمة 0 حتى ولو كانت العلاقة موجودة بين المتغيرين. (قد تكون واضحة بمجرد النظر إلى جدول التقاطع) إن سبب هذه المشكلة يكمن في أن البيانات تنحرف بشكل كبير على طول المتغير التابع (إن قيمة لامبيدا ستكون 0 عندما تكون فئة النموذج للمتغير التابع مساوية لكل الفئات المتعلقة بالمتغير المستقل).

ولكي نرى ذلك عملياً، سوف نقوم بتحليل البيانات الواردة في الجدول التالي (7 - 8):

في هذا المثال النظري قد تم طرح السؤال المتعلق بما إذا كانت دائرة التشغيل تقوم بجهد كبير للتخفيف من مشكلة البطالة.

جدول (7-8):

هل يجب على دائرة التشغيل أن تبذل جهداً أكبر للتخفيف من البطالة

يوافق	العمر		المجموع
	أقل من 45	45 فما فوق	
لا	110 % 18	168 % 42	278
نعم	490 % 82	232 % 58	722
المجموع	600	400	1000

وبالنظر إلى جدول التقاطع فإنه يمكننا القول إنه توجد علاقة محدودة. فكلما زادت نسبة الأفراد الذين تكون أعمارهم تحت 45 سنة يوافقون على السؤال المطروح حول سياسة التشغيل أكثر من أولئك الذين تكون أعمارهم 45 سنة فأكثر. وبوضوح، يمكننا أن نرى أن هناك بعض الاعتماد بين المتغيرين، والذي يمكننا أن نصفه بتعبيرات لفظية بالقول إن هناك علاقة جيدة بعض الشيء إلى متوسطة في القوة.

وعليه إذا أردنا أن نحاول تكميم هذه العلاقة باستخدام لامبيدا فإننا نحصل على تطابق متناسب مساوياً لـِ 0. (لاحظ الإجابة للمتغير التابع لكل الحالات 1000 هي "نعم" والتي هي أيضاً كذلك الإجابة لكل من الفئتين للمتغير المستقل: فغالبية الناس تحت الـِ 45 سنة أجابوا بنعم، وأن الغالبية من الناس ذوي الأعمار 45 سنة فما فوق قد أجابوا هم أيضاً بنعم).

إن هذا التوزيع المنحرف في إطار المتغير التابع سوف يولد لنا لامبيدا تساوي 0، حتى كان واضحاً لدينا من خلال العين المجردة أن هناك وجوداً لبعض التطابق بين المتغيرين. ولكي نبين ذلك فإننا نحتاج أولاً: إلى حساب عدد الأخطاء عند التكهن بدون معلومات حول المتغير المستقل (العمر)، إننا نتوقع أن كل حالات الـِ 1000 سوف تقع

في فئة "نعم"، حيث إن هذا الأمر سوف يقلل معدل الخطأ حيث يصل معدل الخطأ في هذه الحالة إلى 278 خطأ: $E_1 = 278$.

أما إذا كانت لدينا معلومات حول المتغير التابع فإننا سوف نبقى على نفس الأخطاء. فإذا ما نظرنا إلى إجابة المبحوثين الأولى أقل من 45 سنة فإننا نتوقع أن 600 حالة قد أجابت "بنعم" (110خطأ).

ثانياً: يمكننا التوقع أن 400 شخص الذين تبلغ أعمارهم 45 سنة فما فوق قد أجابوا "بنعم" (168 خطأ).

وهذا المجموع الذي يصل إلى 278 خطأ هو نفس التوقعات بدون معرفة عمر المبحوث: $E_2 = 278$. وأن قيمة لامبيدا ستكون:

$$\lambda = \frac{E_1 - E_2}{E_1} = \frac{278 - 278}{278}$$

$$= 0$$

ومن هنا نجد أن لامبيدا قد فشلت في الحصول على علاقة مشاهدة والتي كانت واضحة لنا بالعين المجردة. إن هذا الأمر قد يلقي بعض الضوء على قاعدة مهمة: متى كانت لامبيدا مساوية لـِ 0 ينبغي معاينة التوزيع التكراري النسبي لنقرر فيما إذا كانت هذه التكرارات تعكس في الواقع لا علاقة أو فيما إذا كانت هذه العلاقة جاءت نتيجة لانحراف التوزيع المتعلق بالمتغير التابع. إذا فحصنا نسب العمود تقودنا إلى الاستنتاج أن قيمة 0 للامبيدا هي ناتجة من توزيع منحرف (كما في هذه الحالة) توجد لدينا ثلاثة خيارات:

1- ينبغي على الباحث ألا يشغل باله بمقاييس التطابق وعليه أن يتقيد بجدول التطابق وما يحتويه من توزيعات نسبية، وأن يكون الأساس في نتيجته هو النظر إلى العلاقة في ذاتها فقط. ويتطلب هذا الأمر من الباحث أن يقوم ببعض الأحكام الذاتية، ولكن طالما أن جدول التقاطع يسمح للقراء أن يقيموا بأنفسهم، فإن ذلك لا يشكل أي مشكلة في بناء هذه المحاولة باستخدام التوزيعات النسبية كشواهد. إن هذه

التوزيعات في بعض الأحيان "تتحدث بنفسها". فحساب إحصاءات متقدمة أكثر قد يؤدي إلى اختفاء معلومات مهمة في كتلة من الأعداد المشكوك فيها.

2- ينبغي على الباحث حساب مقاييس أخرى للتطابق حيث توجد مجموعة أخرى من مقاييس التطابق للبيانات المقاسة على المستوى الاسمي والتي يمكن استخدامها إذا ما واجهت الباحث بعض المشكلات المتعلقة بلامبيدا مثل مقاييس جودمان – كروشكل تاو Goodman-Kruskal tau. إن هذا المقياس يشبه مقياس لامبيدا في كونه مقياس لا تماثل للتطابق حيث تتراوح درجاته بين 0 و 1 ولكنه في الجانب الآخر لا يشبه لامبيدا لكونه لا يستخدم إجابة النموذج للمتغير المستقل عند إجراء عملية التنبؤات، ولكنه على العكس يستخدم بدلاً من ذلك التوزيع التكراري للحالات عبر كل الفئات للمتغير المستقل. ولما كان هذا الأمر أقل حساسية لانحراف التوزيعات الهامشية من لامبيدا بالتالي فهو خيار ملائم عندما تسبب عملية الانحراف قيمة لامبيدا المساوية لصفر.

ومن المقاييس الأخرى للتطابق هو Cramer`s V. وهذا المقياس دائماً يولد قيمة أكبر من 0 عندما يكون هناك تطابق بين المتغيرين. إلا أنه في الواقع لا يمتلك تفسيرات بسيطة فيما يتعلق بنسبة التخفيض في الخطأ (PRE).

وعليه فإنه لا يمكن استخدامه لتقييم قوة العلاقة في أي جدول تقاطع. إلا أنه يمكن أن يكون مفيداً عند مقارنة قوة العلاقات الثنائية عبر مجموعة من الجداول. والمعادلة المستخدمة لـِ Cramer`s V هي:

$$V = \sqrt{\frac{X^2}{N(K-1)}}$$

حيث إنّ:

X^2 هي إحصاءات مربع كاي لجدول التقاطع.

K تشير إلى عدد الصفوف أو إلى عدد الأعمدة أيهما أصغر. إِنّ Crame`s V هي إحدى الخيارات الموازية مع لامبيدا، للبيانات الاسمية.

3- إن الخيار الثالث يكمن في معايرة الجدول حتى يكون مجموع الصفوف كلها متساوية. إن مثل هذا الإجراء إجراء معقد إلى حد ما، فعملية المعايرة في الجدول تتضمن محاولة التقليل من التباين في البيانات التي جاءت من خلال التوزيع المنحرف في المتغير التابع، بينما لا يزال الاحتفاظ بالتباين عبر فئات المتغير المستقل. وفي تقرير يستخدم هذا الإجراء ينبغي أن يكون واضحاً أن لامبيدا لا يمكن حسابها ببيانات الصف، وبحساب لامبيدا فإن هوامش الصفوف قد عُيّرت ليصل مجموعها إلى 100. وعند التعامل مع لامبيدا فإننا نقنن هوامش الصف بحيث يصل مجموع كل صف إلى 100. وهذا الأمر يتطلب حساب نسب الصف، التي يمكن التعامل معها كما لو كانت هي الأعداد الحقيقية للحالات. هذا يعني، أننا قمنا بحساب نسبة مجموع "نعم" للمبحوثين الذين هم تحت الـ 45 سنة، وحساب نسبة أولئك الذين تصل أعمارهم إلى 45 سنة أو تزيد. ونقوم بعمل الشيء نفسه لإجابة أولئك الذين أجابوا "بلا". انظر جدول (7 – 9).

جدول (7-9):

هل يجب على دائرة التشغيل أن تبذل جهدا كبيرا للتخفيف من البطالة

يوافق	العمر		المجموع
	أقل من 45	45 فما فوق	
لا	$\frac{110}{278} \times 100 = \%40$	$\frac{168}{278} \times 100 = \%60$	% 100
نعم	$\frac{490}{722} \times 100 = \%68$	$\frac{232}{722} \times 100 = \%32$	% 100

وبعد ذلك يمكننا استخدام أعداد النسب كما لو حسبت كحالات فعلية كما في الجدول التالي:

جدول (7ـ10):

هل يجب على دائرة التشغيل أن تبذل جهداً كبيراً في التقليل من البطالة

يوافق	العمر		المجموع
	أقل من 45	45 فأكثر	
لا	40	60	100
نعم	68	32	100

تذكر أن هذه النسب: 40 تمثل 40 % من 278 لمجموع إجابات "لا". وهكذا. إلا أننا قد تعاملنا مع هذه النسب كما لو كانت حالات فردية. وهذا يعني أن الحجم الكلي للعينة هو 200 بدلاً من 1000: الذين أجابوا "بنعم" يصل عددهم إلى 100 والذين أجابوا "بلا" يصل عددهم إلى 100. وباستخدامنا لهذه البيانات من جدول المعايرة (10)، يمكننا إعادة حساب لامبيدا بدون معرفة حول المتغير المستقل، فيمكننا تصنيف كل الإجابات الـ 200 إما "بنعم" أو "بلا"، وعليه يمكننا أن نتسبب في أخطاء تصل إلى 100.

$$E_1 = 100$$

أما إذا كانت لدينا معرفة حول المتغير المستقل فإنه يمكن أن نصل إلى التوقعات التالية، فإذا بدأنا بأولئك الأشخاص الذين تقل أعمارهم عن 45 سنة، يمكننا أن نتنبأ بأولئك الذين أجابوا "بنعم"، ولما كان هذا التنبؤ يعطينا أقل معدل في الخطأ (40 خطأ) ولأولئك الذين تصل أعمارهم 45 سنة فما فوق، يمكننا التكهن بأنّ كل هؤلاء قد أجابوا "بلا"، وعليه فإننا بذلك قد ولدنا 32 خطأ:

$$E_2 = 40 + 32 + 72$$

ومن هنا تكون قيمة لامبيدا مساوية لـِ:

$$\lambda = \frac{E_1 - E_2}{E_1} = \frac{100 - 72}{100}$$

$$= 0.28$$

وعندما نقرب هذه القيمة (0.28) في المائة (100)، فإن قيمة لامبيدا تشير إلى تطابق ضعيف إلى متوسط، فيما يتعلق بنسبة التخفيض في الخطأ (PRE). وعليه، فإنه يمكن تفسير لامبيدا بالإقرار بأن معرفتنا بالعمر تساعد قدرتنا على أن نتنبأ بالاتجاه حول الجهد المبذول من دائرة التشغيل للتخفيف من البطالة بنسبة 28 %.

أسئلة للمراجعة:

1- بين الفرق بين لا تماثل Asymmetric و متماثل Symmetric، وما هي أنسب هذه المقاييس للاستخدام في موقف يكون فيه متغيران تبادلياً معتمدان على بعضهما البعض Mutually dependent؟

2- ما هو الشيء المهم عندما تقوم بحساب لامبيدا لنقرر ما إذا كان أحد المتغيرين يرجح أن يكون معتمداً على المتغير الآخر، وإذا كان الأمر كذلك، حدد ما هو المتغير التابع وما هو المتغير المستقل؟

3- احسب لامبيدا من البيانات التالية، وبين قوة أي علاقة؟

أ:

تابع	مستقل		المجموع
	1	2	
1	30	60	90
2	45	50	95
مجـ	75	110	185

ب:

تابع	مستقل			المجموع
	1	2	3	
1	56	40	10	106
2	15	30	50	95
مجـ	71	70	60	201

ج:

تابع	مستقل			المجموع
	1	2	3	
1	70	40	10	106
2	50	45	38	133
3	43	30	14	87
مجـ	163	115	62	326

4- يرغب باحث في معرفة الاتجاه نحو التعليم المختلط. وقد تم بحث 3000 طالبٍ وطالبةٍ، وجاءت النتيجة كالتالي:

الاتجاه	النوع		المجموع
	ذكور	إناث	
أوافق	849	367	
لا أوافق	191	1593	

احسب قيمة لامبيدا وفسرها.

5- احسب قيمة Cramerś V للبيانات التالية:

أ- $X^2 = 3.5, N = 20$

2 = Culmns (الأعمدة)، 4 = rows (الصفوف)

Rows = 2, Culmns, $X^2 = 9.8, N = 90 = 4$

Rows = 3, Culmns, $X^2 = 12, N = 800 = 3$

6- مسح لعدد 50 طالبةً من الإناث و50 طالباً من الذكور، طرح عليهم سؤال هل تعرف تاريخ وضع حجر الأساس لمشروع النهر الصناعي العظيم؟

الذكور أجابوا "بنعم" 29

الإناث أجابوا "بلا" 21

الإناث أجابوا "بنعم" 22

الذكور أجابوا "بلا" 28

- رتب هذه البيانات في جدول تقاطع Crosstabs. وبين المتغير المستقل والمتغير التابع.
- احسب قيمة لامبيدا لهذه البيانات؟

7- دراسة توصلت إلى نتيجة مفادها أن التطابق بين متغيرين باستخدام Cramerś V، 0.34. وفي دراسة سابقة قامت بقياس التطابق بين نفس هذين المتغيرين باستخدام V كانت النتيجة تتراوح بين 0.15 إلى 0.21. كيف لهؤلاء الباحثين تفسير هذه النتائج؟

الهوامش والمصادر:

أولا: الهوامش:

1- George Argyrous, Statistics for Social and Health Research With a Guide to SPSS, Sage Publications ,London, 2001, P.153.

2- Ibid, PP. 154 - 155.

3- Ibid, PP. 159 - 161.

4- Ibid, P. 163.

5- George Argyrous, opt. Cit., PP. 165 - 196.

6- And, Joseph F. Healey, The Essentials of Statistics: A Tool for Social Research, Wadsworth Cengage Learning, USA, 2010. pp. 295 .

ثانيا: المصادر:

1- George Argyrous, Statistics for Social and Health Research With Guide to SPSS, Sage Publications, London, 2001 .

2- Joseph F. Healey, The Essentials of Statistitics: A Tool for Social Research, Wadsworth Cengate Learning, USA. 2010.

3- William E. Wagner, III, using SPSS for Social Statistics and Research Methods 2nd ed, Pine forge Press An Inprint of Sage, USA, 2010.

الفصل الثامن

التطابق بين متغيرات تم قياسها على المستوى الترتيبي

الوصف العددي للبيانات الترتيبية

مقدمة

نحاول في هذا الفصل التركيز على مقاييس التطابق لمتغيرات تم قياسها على المستوى الترتيبي لنسبة التخفيض من الخطأ (PRE) التي هي مشابهة لمقياس لامبيدا من حيث المنطق الأساسي لهذه المقاييس والكيفية التي يتم بها تفسير البيانات. وعند حديثنا عن لامبيدا فإننا نحاول من خلالها التنبؤ بقيمة حالة فردية أخذت للمتغير المستقل. والهدف من وراء ذلك هو أولاً: افتراض أنه لا يوجد تطابق بين المتغيرين. وثانياً: بافتراض أنه يوجد تطابق تام بين المتغيرين. وبمقارنة معدلات الخطأ المتعلق بكل قاعدة من قواعد التنبؤ فإننا نستطيع تقييم العلاقة الفعلية التي تحتويها مجموعة البيانات التي تم جمعها.

وتجدر الإشارة إلى أننا سنتبع نفس الإجراء مع البيانات الترتيبية، مع الأخذ في الاعتبار أننا سوف نستخدم بيانات إضافية حول المتغيرات المعطاة للتعامل معها وفقاً لمستوى البيانات الترتيبية، تختلف البيانات الترتيبية عن البيانات الاسمية، غير أننا في البيانات الترتيبية على دراية بالكيفية التي رتبت بها الحالات. وعليه فإننا نحاول أن نتنبأ

بالوضع الترتيبي لحالات الأزواج Pairs of Cases، وقاعدة مقاييس التطابق على أساس النجاح في تنبئنا لهذه الرتب. فإذا ما تعاملنا مع المثال الموضح في جدول (1) فإنه بإمكاننا أن نستمر في تكميم العلاقة القوية الموجبة ونلاحظ من خلال حساب مقاييس التطابق للبيانات الترتيبية ذات الصلة أنه توجد لدينا مجموعة من مقاييس التطابق المتعلقة بنسبة التخفيض في الخطأ التي يمكن حسابها من هذا الجدول مع تباين بسيط في طرقها الخاصة؛ وبذلك فإن كل المقاييس المتعلقة بالبيانات الترتيبية والتي سوف نناقشها لديها خصائص مشتركة تستند على التميِّز بين الأزواج المتوافقة Concordant Pairs والأزواج غير المتوافقة Disconcordant Pairs.

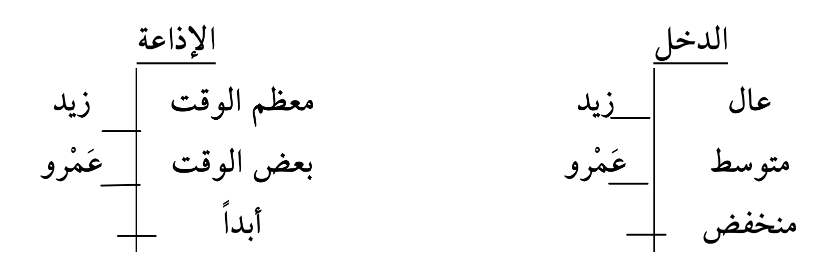
جدول (8-1): توزيع مشاهدة الإذاعة المرئية وفقا للدخل

مشاهدة الإذاعة المرئية	الدخل			المجموع
	منخفض	متوسط	عال	
أبداً	75 %71.5	15 %15.0	10 %10.5	100
بعض الوقت	20 %19.0	70 %70.0	10 %10.5	100
معظم الوقت	10 %9.5	15 %15.0	75 %79.0	100
المجموع	**105**	**100**	**95**	**300**

المصدر: George Argyrous, op.cit, p. 173

1- الأزواج المتوافقة:

نفترض أن واحداً من 75 في فئة الدخل العالي في جدول (1) يشاهد الإذاعة المرئية معظم الوقت ويدعى زيداً، وأن واحد من 70 في فئة الدخل المتوسط يشاهد الإذاعة المرئية بعض الوقت ويدعى عَمْراً. إن هذين الشخصين يمكن ترتيبهما بحيث يخالف أحدهما الآخر على كل واحد من هذين المتغيرين. انظر شكل (8 – 1).



شكل (8-1) ترتيب الزوج المتوافق

إن ترتيب زوج هاتين الحالتين يمكن تلخيصه في الجدول التالي:

جدول (8-2): حالات الزوجين المتوافقين

المتغير المستقل: الدخل	المتغير التابع: الإذاعة المرئية
زيد رتب فوق عَمْرو (لديه أعلى دخل)	زيد رتب أعلى من عَمْرو (مشاهدة أكثر للإذاعة المرئية)

وعليه فإن هاتين الحالتين قد تم ترتيبهما بشكل متساوٍ على كل واحد من هذين المتغيرين. وهذا قد يبدو غريباً في طرحه: حيث يمكن طرح السؤال التالي وهو: كيف يمكن أن يكون ترتيبهما متساوٍ إذا كان لكل منهما قيم مختلفة؟ إن القضية هي أن ترتيبهما متساوياً: إن زيداً قد رتب فوق عَمْرو على كلا المتغيرين. ونصف أزواج مثل هذه الحالات كزوج متوافق (Nc). ونعني بالزوج المتوافق: هو ذلك الزوج الذي يتشكل من خلال حالتين في توزيع متحد الذي رتب بشكل متساوٍ على كلا المتغيرين.

لقد أخذنا حالتين من كل الحالات البالغ عددها 300 والتي تشكل الزوج المتوافق. إن السؤال المطروح هو: هل نقوم بحساب العدد الكلي للأزواج المتوافقة التي يحتويها الجدول؟ ولفعل ذلك انظر إلى الخلايا المظَلَّلة في جدول التوافق التي من خلالها تم سحب كل من زيد وعَمْرو.

جدول (8-3) توزيع مشاهدة الإذاعة المرئية وفقا للدخل

مشاهدة الإذاعة المرئية	الدخل		
	منخفض	متوسط	عال
أبداً	75	15	10
بعض الوقت	20	70	10
معظم الوقت	10	15	75

في مناقشتنا أعلاه قد تم صياغة الزوج المتوافق بمماثلة زيد الذي هو واحد من الحالات التي تبلغ 75 حالة ولديه دخل عالٍ ويشاهد الإذاعة المرئية معظم الوقت. بعَمْرو الذي هو واحد من 70 حالة ذات الدخول المتوسطة ويشاهد الإذاعة المرئية بعض الوقت.

في واقع الأمر يمكننا أن نُقْرِنَ عَمْراً مع كل واحد وأي واحد من 70 حالة في الخلية التي تحتوي على دخل متوسط / ومشاهدة الإذاعة بعض الوقت يولد لنا 70 حالة من الأزواج المتوافقة: عَمْرو مضافاً إلى كل حالة من السبعين حالة الموجودة في الخلية الوسطى من الجدول (شاملة عَمْراً). وعندئذٍ يمكننا عمل الشيء نفسه لكل واحد من 74 حالة التي لديها دخل عالٍ وتشاهد الإذاعة المرئية معظم الوقت. إن هذا سوف يعطينا في المجموع النهائي 75 حصة في 70 أزواج متوافقة: $75 \times 70 = 5250$.

وبالنظر إلى جدول (8 – 4) يمكننا أن نرى أن هناك مجموعات متحدة التي بدورها تشكل الأزواج المتوافقة.

جدول (8-4): توزيع مشاهدة الإذاعة المرئية وفقاً للدخل

مشاهدة الإذاعة المرئية	الدخل		
	منخفض	متوسط	عال
أبداً	75	15	10
بعض الوقت	20	70	10
معظم الوقت	10	15	75

إن كل واحد من 75 حالة في أدنى الخلية اليسرى أيضاً تم ترتيبه فوق كل حالة من الحالات 15 في خلية أبداً / الدخل المتوسط كلاهما لديه دخل عالٍ ويشاهد الإذاعة المرئية أكثر. وبهذا سوف نضيف الأرقام التالية للأزواج المتوافقة:

$$75 \times 15 = 1125$$

في الحقيقة أن أي حالة سوف تشكل زوجاً متوافقاً مع أي حالة أخرى في الخلية التي هي فوق وإلى اليسار منها في الجدول وبالتالي فإن العدد الإجمالي للأزواج المتوافقة سيكون مبيناً في الجدول (8 – 5).

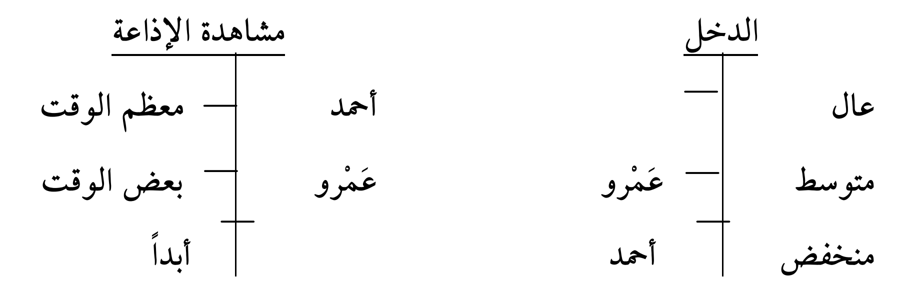
جدول (8-5): حساب الأزواج المتوافقة

75	15	10	
20	70	10	(75×70) +75×15) +75×20) + (75×75 = 13.500
10	15	75	
75	15	10	
20	70	10	(10 × 15) + (10 × 75) = 900
10	15	75	
75	15	10	
20	70	10	(15 × 20) + (15 × 75) = 1425
10	15	75	
75	15	10	
20	70	10	(70 × 75) = 5250
10	15	75	

$$N_C = 13.500 + 900 + 1425 + 5250 = 21.075$$

2- الأزواج غير المتوافقة :

دعنا الآن نأخذ واحداً من عشرة أشخاص الذين لديهم دخل منخفض ويشاهدون الإذاعة المرئية معظم الوقت، ويدعى أحمد، ومقارنته بعَمْرٍو (هو واحد من 70 حالة ذوي الدخول المتوسطة ويشاهدون الإذاعة المرئية بعض الوقت)، فإن الترتيب لن يكون متساوياً لكلا المتغيرين فأحمد تم ترتيبه تحت عَمْرٍو فيما يتعلق بالدخل، ولكنه على الطرف الآخر رتب فوق عَمْرٍو فيما يتعلق بمشاهدة الإذاعة المرئية، انظر شكل (8 – 2). إن مثل هذه الحالات نطلق عليها الأزواج غير المتماثلة (Nd). ونعني بالزوج غير المتماثل هو ذلك الزوج الذي يتشكل من خلال حالتين في توزيع متحد الذي يرتب فيه متغير واحد بشكل مختلف في ترتيبه للمتغير الآخر.



شكل (8-2) ترتيب الزوج غير المتوافق

هذه الحالة سوف تشكل زوجاً غير متوافق مع الحالة الأخرى في الجدول أي في أي خلية أعلى وإلى اليمين. ولحساب العدد الإجمالي للأزواج غير المتوافقة ينبغي علينا أن نبدأ بالخلية اليسرى أسفل الجدول (8 – 6) وتماثلها مع كل الخلايا فوق وإلى اليمين منها.

جدول (8-6): حساب الأزواج غير المتوافقة

75	15	10	
20	70	10	$(10\times70) + (10\times15) + (10\times10) + (10\times10) = 1050$
10	15	75	
75	15	10	
20	70	10	$(20 \times 15) + (20 \times 10) = 500$
10	15	75	
75	15	10	
20	70	10	$(15 \times 10) + (15 \times 10) = 300$
10	15	75	
75	15	10	
20	70	10	$(70 \times 10) = 700$
10	15	75	

$$N_d = 1050 + 500 + 300 + 700 = 5220$$

مقاييس التطابق للمتغيرات الترتيبية:

كل نسب التخفيض في الخطأ في مقاييس التطابق (PRE) تستخدم الفرق بين الأزواج المتوافقة والأزواج غير المتوافقة كأساس لتقييم ما إذا كان هناك تطابق وتحديد اتجاه هذا التطابق. إن السبب وراء النظر حول هذه الأزواج المتوافقة وغير المتوافقة أنها تعطينا معلومات يمكننا استخدامها في التنبؤ بما إذا كان لدينا متغيران متطابقان بشكل

موجب بعدئذٍ فإن جدول التقاطع سوف يحتوي أكثر الأزواج المتوافقة منه من الأزواج غير المتوافقة، والعكس بالعكس عندما يكون لدينا تطابق سالب.

وتجدر الإشارة إلى أنه عندما يكون هناك تطابق موجب بين المتغيرين فإن البيانات سوف تحتوي كثيراً من الأزواج المتوافقة وقليلاً من الأزواج غير المتوافقة. وإذا كان الأمر كذلك، فإن علمنا بأن شخصاً ما ترتيبه فوق شخص آخر فيما يتعلق بالدخل عندئذٍ يمكننا أن نتوقع أن ذاك الشخص يحتل كذلك مرتبة أعلى فوق الشخص الآخر فيما يتعلق بتوزيع مشاهدة الإذاعة المرئية.

التطابق الموجب: $N_c - N_d > 0$

أما في حالة العلاقة السالبة بين المتغيرين فإن البيانات سوف تحتوي على كثير من الأزواج غير المتوافقة ومن هنا يمكننا الوصول إلى تكهن عكسي: فمعرفتنا بشخص رتب فوق شخص آخر فيما يتعلق بالدخل سيقودنا إلى أن نتكهن بأن هذا الشخص يحتل مرتبة أدنى من الشخص الآخر فيما يتعلق بتوزيعات مشاهدة الإذاعة المرئية:

التطابق السالب: $N_c - N_d < 0$

أما فيما يتعلق بعدم وجود تطابق بين المتغيرين فإن البيانات ستحتوي تماماً على كثير من الأزواج المتوافقة والأزواج غير المتوافقة ومن هنا نكون غير قادرين على زيادة قدرتنا في التكهن بفئة المتغير التابع من خلال معرفتنا بفئة المتغير المستقل التي تقع فيها هذه الفئة:

لا تطابق: $N_c - N_d = 0$

وتجدر الإشارة هنا إلى أن هناك أربعة أسس لنسبة التخفيض في الخطأ في مقاييس التطابق للبيانات الترتيبية: سومرز Somerś d d وجاما Gamma وكنلدتاو Kendalś tau-b b وكنلدتاو Kendalś tau-c C. كل هذه المقاييس متشابهة في كونها تعطينا تفسيرات لنسبة التخفيض في الخطأ، كما أن كل هذه المقاييس تستخدم الفرق بين Nc و Nd كأساس لتقييم قوة العلاقة. إلا أن الفرق بين هذه المقاييس يكمن في كيف يمكننا معايرة هذا الفرق. ولمناقشة هذه المقاييس دعنا نبدأ بأبسطها المتمثل في جاما.

جاما Gamma:

تعتبر جاما من المقاييس الشائعة لـِ (PRE) لقياس التطابق بين متغيرين تم قياسهما على الأقل على المستوى الترتيبي وتم ترتيب هذين المتغيرين في جدول ثنائي. كما تعتبر جاما مقياساً متماثلاً للتطابق. وعليه فإن القيمة المحسوبة ستكون واحدة بغض النظر عن الكيفية التي يتحدد فيها المتغير المستقل والطريقة التي يحدد بها المتغير التابع. بمعنى آخر، إذا قلبنا الصفوف والأعمدة في الجدول بمعنى أن يكون الدخل تحت عند الصفوف ومشاهدة الإذاعة المرئية يكون عبر الأعمدة. فإن حساب جاما لن يتأثر. وعليه فإن جاما لن تكون حساسة لنموذج معين نعتقد أنه يصف العلاقة بين المتغيرين.

إن معادلة جاما تبين لنا الفرق بين عدد الأزواج المتوافقة والأزواج غير المتوافقة كنسبة من العدد الكلي للأزواج المتوافقة والأزواج غير المتوافقة وباستخدام البيانات من المثال السابق فإن حساب جاما يكون كالتالي:

$$G = \frac{N_C - N_d}{N_C + N_d} = \frac{21.075 - 2550}{21.075 + 2550}$$

$$= 0.78$$

وتشير هذه النتيجة إلى تطابق قوي موجب بين هذين المتغيرين، تلك النتيجة التي تعزز النتيجة التي وصلنا إليها استناداً إلى التحليل النظري لجدول التطابق. إنه من الوضوح أن المدى الممكن لقيم جاما تتراوح بين 1− و +1. فعندما تكون جاما 1− فإنها تشير إلى تطابق تام سالب: فمعرفة أن الحالة مرتبة فوق الحالة الأخرى على واحد من المتغيرين تشير إلى أنه يجب أن يرتب تحت المتغير الآخر. إن مثل هذه النتيجة نتحصل عليها إذا كان لدينا فقط الأزواج غير المتوافقة كما يوضحه الجدول التالي:

جدول (8-7)

توزيعات مشاهدة الإذاعة المرئية وفقا للدخل: تطابق تام سالب

مشاهدة الإذاعة المرئية	الدخل		
	منخفض	متوسط	عال
أبداً	0	0	%100
بعض الأوقات	0	%100	0
كل الأوقات	%100	0	0

أما على الجانب الآخر، إذا كان لدينا فقط أزواج متماثلة فإن قيمة جاما ستكون +1 مشيرة إلى تطابق موجب تام يعني: معرفتنا بترتيب حالة فوق الحالة الأخرى على المتغير المستقل تشير إلى أن هذه الحالة يجب أيضاً أن تكون مرتبة فوق المتغير التابع. إن مثل هذا الموقف ينعكس في الجدول التالي:

جدول (8-8)

توزيعات مشاهدة الإذاعة المرئية وفقا للدخل: تطابق تام موجب

مشاهدة الإذاعة المرئية	الدخل		
	منخفض	متوسط	عال
أبداً	%100	0	0
بعض الأوقات	0	%100	0
معظم الأوقات	0	0	%100

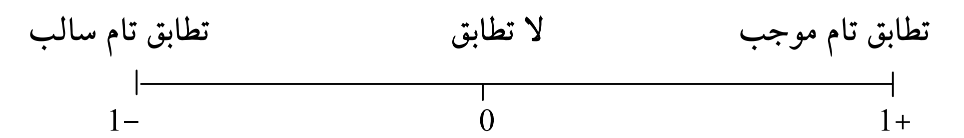
عندما تكون قيمة جاما صفر فهي تشير إلى عدم وجود تطابق. فإذا كان هناك فقط مجموعة كبيرة من الأزواج المتماثلة مثل وجود مجموعة كبيرة من الأزواج غير المتماثلة، عندئذٍ فإن معرفة الترتيب على طول متغير واحد لا يعطينا أي دليل في كيفية ترتيبه على المتغير الآخر. ويتضح هذا الموقف جلياً في الجدول التالي:

جدول (8-9)

توزيعات مشاهدة الإذاعة المرئية وفقا للدخل: لا تطابق

مشاهدة الإذاعة المرئية	الدخل		
	منخفض	متوسط	عال
أبداً	%50	0	%50
بعض الأوقات	0	%100	0
معظم الأوقات	%50	0	%50

تشير الجداول السابقة إلى توضيح النقاط الثلاثة المتطرفة على المقياس المعياري لقياس قوة التطابق بين متغيرين ترتيبيين. كما هو موضح في الشكل التالي:



شكل (8-3) مدى قيمة جاما

وبشكل واضح فإن البيانات الواردة في هذا المثال لا تتطابق مع أي واحد من هذه المواقف الثلاثة المتطرفة. إذاً السؤال الذي يطرح الآن هو أي من قواعد التنبؤ ستكون قريبة من النتائج التي تحصلنا عليها فعلياً؟ إنه من الوضوح أن جدول التطابق التام الموجب يمثل واحدة من تلك البيانات الفعلية الأقرب تشابهاً. فقيم جاما التي تحصلنا عليها (0.78) لا نجدها تماماً 1+ ولكنها قريبة منه أكثر منه إلى 0 أو -1.

وتمتاز جاما بشيوعية استخدامها نتيجة لسهولة حسابها نسبياً، إلا أن هذه الميزة قد أبطلت باستخدام برنامج العقل الآلي مثل (SPSS) الذي من خلاله يمكن حساب كل المقاييس بكل سهولة ويسر. إلا أنه قد يساورنا الشك أن عنصراً آخر قد يكون مرتبطاً بشيوعية استخدام جاما ولديه القدرة على توليد قيمة عالية لقوة التطابق مقارنة بمقاييس

تطابق أخرى ترتيبية. وتجدر الإشارة إلى القول بأن جاما تعتريها بعض القيود ينبغي علينا أن نكون على وعي بها:

أولهما: أن مقياس جاما هو فقط مقياس متماثل، وعليه لا يأخذ ميزة المعلومات المقدمة من خلال البيانات التي نعتقد أنها نموذج أكثر ملاءمةً لوصف العلاقة التي تكون معتمدة على طريقة واحدة ويتمثل هذا القيد الرئيسي الآخر عندما يكون التطابق تاماً سوف تولد لنا جاما قيمة +1 أو –1 والعكس ليس دائماً صحيحاً: فقيمة جاما + أو –1 لا تشير دائماً إلى تطابق تام. فقد يكون بإمكاننا أن نولد قيمة لجاما –1 أو +1 لجدول تقاطع حتى عندما يكون هناك بشكل واضح أقل من التطابق التام في البيانات. ويحدث هذا الأمر عندما تكون العلاقة غير مُتسقة. هنا ينبغي علينا اتباع القاعدة، وعليه فإنه قبل استخدام جاما لبيانات جدول ثنائي ينبغي علينا فحصها لتقييم ما إذا كانت العلاقة علاقة مُتسقة. كلا هذين القيدين في حقيقة الأمر ناشئُ من نفس خاصية حساب جاما. وهذا يتضمن عجزاً مرتبطاً بجاما يتضمن الحالات المتعادلة في معادلتهما. وعليه توجد ثلاثة أنماط من الحالات المتعادلة:

1- الحالات المتعادلة على المتغير المستقل (Tx) هذه الأزواج للحالات التي لديها نفس القيمة للمتغير المستقل ولكنها في ذات الوقت لديها قيم مختلفة للمتغير التابع. وهذا في الواقع أن أي حالتين في نفس عمود جدول التقاطع ولكنهما في صفوف مختلفة. في المثال السابق هذه الأزواج من الحالات التي لديها نفس الدخل ولكنها تشاهد كمية مختلفة من الإذاعة المرئية.

2- الحالات المتعادلة على المتغير التابع (Ty) فإن هذه الأزواج للحالات لديها نفس القيمة للمتغير التابع ولكنها في ذات الوقت لديها قيم مختلفة للمتغير المستقل. ومن الناحية العملية، فإن أي حالتين في نفس الصف بجدول التقاطع ولكنها في أعمدة مختلفة. في المثال الذي بين أيدينا هذه الأزواج من الحالات تشاهد نفس كمية الوقت ولكن لديها دخل منخفض.

3- حالات متعادلة على كلا المتغيرين (Tyx) هذه الحالات لديها نفس القيمة في كلا المتغيرين. فهذه أزواج من الحالات سحبت من نفس الخلية في الجدول. ففي المثال

الذي أوردناه فإن هذه الأزواج من الحالات لديها نفس الدخل وتشاهد نفس الكمية من الوقت "الإذاعة المرئية".

وأخيراً توجد مقاييس أخرى متعلقة بنسبة التخفيض في الخطأ في مقاييس التطابق للبيانات الترتيبية التي نسعى إلى تعويض هذا القصور المرتبط بجاما وذلك بإضافة بعض أو كل هذه الحالات المتعادلة في عملية حساباتها [2].

سومرز Somerś d d:

يعتبر مقياس سومرز d مقياساً غير متماثل للتطابق وبذلك يكون هذا المقياس ذا حساسية لأي متغير يوصف بأنه متغير مستقل أو يوصف آخر بأنه متغير تابع. وعليه فإنه من المفيد عندما نشعر بالعلاقة بين متغيرين أن أفضل وصف لهذه العلاقة يتوقف على نموذج اعتماد الاتجاه الواحد. إن المنطق وراء استخدام سومرز d يعتمد في الأساس على فكرة أن الحالتين اللتين تتباينان فيما يتعلق بالمتغير المستقل ولكنهما لا تتباينان فيما يتعلق بالمتغير التابع (هذه الحالات متعادلة على المتغير التابع) تعكس لا تطابق. في المثال الذي تعاملنا معه، فالأزواج المتعادلة على المتغير التابع ولكنها ليست كذلك على المتغير المستقل فإن تلك الأزواج من الحالات تختلف فيما يتعلق بالدخل ولكن مشاهدة التطابق تماماً نفس كمية الإذاعة المرئية. وتحسب قيمة سومرز d كنسبة كل الحالات المتوافقة وغير المتوافقة مضافاً إليها الأزواج المتعادلة على المتغير التابع:

$$d = \frac{N_C - N_d}{N_C + N_d + T_y}$$

ولحساب عدد الحالات المتعادلة فإننا نأخذ كل خلية، بَدْءًا من أعلى اليسار ونضربها في عدد الحالات التي تحتويها في عدد الحالات في الخلايا إلى اليمين منها. انظر جدول (8 – 10) وبتعويض هذه الحسابات في معادلة سومرز d نحصل على القيمة التالية:

$$d = \frac{N_C - N_d}{N_C + N_d + T_y} = \frac{21.075 - 2550}{21.075 + 2550 + 6350}$$

$$= 0.62$$

إن هذه القيمة 0.62 تشير إلى تطابق موجب، متوسط بين هذين المتغيرين فالزيادة في الدخل مرتبطة بالزيادة في مشاهدة الإذاعة المرئية.

جدول (8-10) حساب الحالات المتعادلة على المتغير التابع

75	15	10	
20	70	10	$(75\times15) + (75\times10) = 1875$
10	15	75	
75	15	10	
20	70	10	$(15 \times 10) = 150$
10	15	75	
75	15	10	
20	70	10	$(20 \times 70) + (20 \times 10) = 1600$
10	15	75	
75	15	10	
20	70	10	$(70 \times 10) = 700$
10	15	75	
75	15	10	
20	70	10	$(10 \times 15) + (10 \times 75) = 900$
10	15	75	
75	15	10	
20	70	10	$(15 \times 75) = 1125$
10	15	75	

$$T_y = 1875 + 150 + 1600 + 700 + 900 + 1125 = 3506$$

لاحظ أن معادلة سومرز d تقريباً تكون مساوية لمعادلة جاما ما عدا أن المقام في معادلة سومرز d لإعداد المتغير التابع تكون متعادلة. وكنتيجة لذلك عندما يكون لدينا مثل هذه الحالات المتعادلة، فإن قيمة سومرز d ستكون دائماً قيمتها أقل من قيمة جاما.

بمعنى آخر، بتجاهل الحالات المتعادلة فإن جاما يمكن أن يبالغ في قوة التطابق بين المتغيرين في علاقة غير متماثلة، لاسيما عندما تكون هناك مجموعة كبيرة من الحالات المتعادلة وحيث إنّ سومرز d هو مقياس غير متماثل للتطابق فإنه بإمكاننا في الواقع حساب خيارين لهذا المقياس، لأي جدول تقاطع. بمعنى آخر، يمكننا حساب سومرز d من خلال متغير واحد كمتغير مستقل والمتغير الآخر كمتغير تابع، كما أنه بإمكاننا قلب هذين المتغيرين بكيفية أخرى وحساب قيمة سومرز d مرة ثانية. ففي المثال السابق قمنا بحساب سومرز d متعامِلِينَ مع الدخل كمتغير مستقل ومشاهدة الإذاعة المرئية كمتغير تابع، وبما أن النموذج النظري لهذه العلاقة يصور لنا أن السببية تسير في ذلك الاتجاه فقد يكون شخص ما لديه نظرية مختلفة التي تفترض بطريقة ما أن مشاهدة الإذاعة المرئية تحدد مستوى الدخل استناداً إلى ذلك يصبح في مقدوره حساب قيمة سومرز d كخيار آخر مستخدماً الدخل كمتغير تابع متحصلاً على قيمة مختلفة لـِ سومرز d.

كندل - تاو Kendalls tau-b: b:

يعتبر كندل تاو b مقياساً للتطابق غير متماثل لنسبة التخفيض في الخطأ (PRE) لبيانات ترتيبية تم ترتيبها في جدول ثنائي. والميزة الرئيسية المرتبطة بمقياس كندل تاو b تكمن في استخدامه للمعلومات التي توفرها الحالات المتعادلة على المتغير التابع والمتغير المستقل:

ومن خلال المبادئ الرياضية يمكننا ملاحظة أن تاو b تمثل الوسط الهندسي للقيم البديلة لسومرز d. وبالرغم من أن هذا المفهوم قد يكون مشوشاً إلى حد ما نظراً لأن سومرز d من خلال التعريف هو مقياس غير متماثل. ولما كانت قيمة تاو b هي الوسط الهندسي لمقياس سومرز d فإن هذه القيمة ستكون في مكان ما بين القيمتين المتعلقتين بسومرز d التي يمكن حسابها من أي جدول تقاطع أمكن استخدامه. وتجدر الإشارة، إلى أن قيمة تاو b يتراوح مداها بين -1 و +1 حيث الأرقام المتعلقة بالصفوف تكون مساوية لعدد الأعمدة. وعليه وبشكل عام فإن تاو b تستخدم في حالات خاصة.

كندل تاو - C Kendallś tau-c:

يعتبر كنلدتاو C مقياساً متماثلاً لنسبة التخفيض في الخطأ للبيانات يشبه كثيراً مقياس تاو b. ويستخدم تاو C في مواقف تكون فيها رغبة الباحث في قياس التماثل في جدول يحتوي على أعداد غير متساوية من الصفوف والأعمدة التي تحتوي على عدد كبير من الحالات المتعادلة. إن المعادلة الدقيقة لِـ tau-c:

$$tau - C = \frac{2k(N_C - N_d)}{N^2(k-1)}$$

حيث إنّ:

K عدد الصفوف أو عدد الأعمدة أيهما أصغر،

N مجموع عدد الحالات.

في المثال السابق لدينا نفس عدد الصفوف مساوياً لعدد الأعمدة (ثلاثة) لذلك فإن

K = 3

وأن قيمة tau-c ستكون:

$$tau - C = \frac{2k(N_C - N_d)}{N^2(k-1)} = \frac{2(3)(21.075 - 2550)}{300^2(3-1)}$$

$$= 0.62$$

ويتضح مما سبق أن هذه النتيجة مساوية لقيمة سومرز d التي تم حسابها فيما سبق ولذلك فإن هذين البديلين من المقاييس تتفق مع بعضها البعض في النتائج [3].

إجراء مقاييس التطابق باستخدام SPSS:

1- من القائمة الموجودة في الجزء العلوي من الشاشة انقر فوق:

Crosstabs ⇐ Statistics ⇐ Descriptive ⇐ Analyze

2- انقر فوق Frequency of TV. Watching.

3- انقر فوق ◄ الذي يشير إلى Row(S) في القائمة المحددة للمتغيرات Variables List.

4- انقر على المتغير الذي يشكل الصفوف في الجدول، في هذه الحالة Income Level.

5- انقر على ◄الذي يشير إلى Column(s) في القائمة المحددة للمتغيرات Variable List ستقود هذه العملية إلى لصق Income Level في Column(s) في القائمة المحددة للمتغيرات Income Level.

6- انقر على المربع الذي يكتب عليه Statistics، فيظهر مربع Statistics box: Crosstabs، وهنا تجد مساحة معنونة بِـ: Ordinal التي توفر قائمة بمقاييس التطابق المتوفرة لهذا المستوى من القياس.

7- يختار الباحث معامل Gamma و Somerś dبالنقر على الصندوق التالي لهما أو القريب إليهما. ثم وضع علامة T على صندوق المعامل المطلوبة للإشارة إلى الإحصاءات المختارة.

8- انقر فوق Continue.

9- أنقر على OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

تفسير مقاييس التطابق من مخرجات SPSS:

يمكن للباحث أن يتحصل على نتائج مقاييس التطابق من مخرجات SPSS. وهذه المقاييس هي جزء من أوامر جدول التقاطع Crosstabs Command.

والشكل التالي يبين مخرجات SPSS لإحصاءات جداول التوافق Crosstabs Statistics.

Directional Measures

	Value	Asymp. Stand. Error[a]	Appro T_b	Appro sig.
Ordinal by Ordinal Sommerś d Symmetric Frequency of TV. Watching	.618	0.043	14.192	.000
Dependent	.618	0.043	14.192	.000
Income Level dependent	.618	0.043	14.192	.000

a. Not Assuming the Null hypothesis.

b. using the Asymptotic standard Error assuming the Null hypothesis.

Symmetric Measures

	Value	Asymp. Stand. Error[a]	Appro T_b	Appro sig.
Ordinal by Ordinal Gamma	.784	0.043	14.192	.000
N. of Valid Cases	300			

a. Not Assuming the Null hypothesis.

b. using the Asymptotic standard Error assuming the Null hypothesis

المصدر: George Argyrous, op.cit, P. 184

شكل (8-4): مخرجات SPSS لجداول التقاطع: مخرجات إحصائية

إن حزمة SPSS تنتج لنا القيم المتعلقة بـِ Gamma، وسومرز d: Sommerś d، في مربعين منفصلين يمثل أحدهما مقياساً متماثلاً والآخر مقياساً غير متماثل (يطلق عليه مقياس مباشر Directional Measures في حزمة SPSS. وفي كلتا الحالتين نجد أن القيمتين يولدهما برنامج SPSS، في العمود المعنون "بالقيمة" Value في كل مربع، فهي نفس القيم التي تم حسابها يدوياً. وتعكس هذه القيم تطابقاً متوسطاً، إلى تطابق قوي بين هذين المتغيرين لهذه الحالات المدروسة. وإذا وجد تطابق سالب، فإن الإشارة ناقص (−) سوف تطبع أمام القيمة.

معامل سبيرمان للرتب [4]:

كما نعرف أن ارتباط بيرسون (r) يقيس درجة العلاقة الخطية Linear relationship بين متغيرين عندما تكون البيانات (قيم X و y) تحتوي على درجات عددية مقاسة على المستوى ذي المسافات أو النسبي، وعلى أية حال، فقد تطورت مُعاملات أخرى للعلاقات غير الخطية NonLinear relationship لأنماط أخرى من البيانات، وأن أحد أهم هذه المقاييس قد أطلق عليه ارتباط سبيرمان. ويستخدم ارتباط سبيرمان في موقفين أولهما تستخدم معامل سبيرمان لقياس العلاقة بين X و y، عندما يكون هذان المتغيران مقاسين على مستوى المقياس الترتيبي، وثانيهما تعتبر معامل سبيرمان للرتب بديلاً ذا قيمة لمعامل بيرسون (r) حتى عندما تكون الدرجات الخام الأصلية مقاسة على المستوى ذي المسافات / والنسبي.

إذاً يستخدم معامل سبيرمان في موقفين هامين هما:

1- عندما تكون البيانات الأصلية بيانات ترتيبية أي عندما ترتب قيم X و y. في هذه الحالة، يمكننا ببساطة تطبيق معادلة ارتباط بيرسون (r) لمجموعة الرتب.

2- يستخدم معامل ارتباط سبيرمان عندما يرغب الباحث في مقياس اتساق العلاقة بين X و y. في هذه الحالة، تحول الدرجات الأصلية أولاً إلى رتب؛ وبعد ذلك تستخدم معادلة معامل بيرسون (r) للرتب، باعتبار أن معادلة بيرسون تقيس الدرجة على

النحو الذي تنطبق فيه الرتب على الخط المستقيم a straight Line، كما أن معادلة بيرسون (r) تقيس درجة الاتساق في العلاقة للدرجات الأصلية، وبالمصادفة، عندما يوجد اتساق في علاقة ذات اتجاه واحد بين المتغيرين فإننا نطلق على هذه العلاقة علاقة التماثل Monotonic. عليه، فإن ارتباط سبيرمان يمكن استخدامه لقياس درجة العلاقة التماثلية بين متغيرين. في أي من الحالتين، فإن ارتباط سبيرمان يمكن بيانه بالرمز r_s لنميزه عن ارتباط بيرسون. إن العملية الكاملة لحساب سبيرمان تحتوي على درجات مرتبة. والمثال التالي يوضح هذه العملية:

إن البيانات التالية تبين العلاقة التماثلية التامة بين X و y. عندما تزيد قيمة X. فإن قيمة y بالتالي تتجه نحو النقصان. ولحساب ارتباط سبيرمان، ينبغي على الباحث بادئ ذي بدء ترتيب القيم المتعلقة بـِ X و y، وبعد ذلك يقوم بحساب ارتباط بيرسون لهذه الرتب.

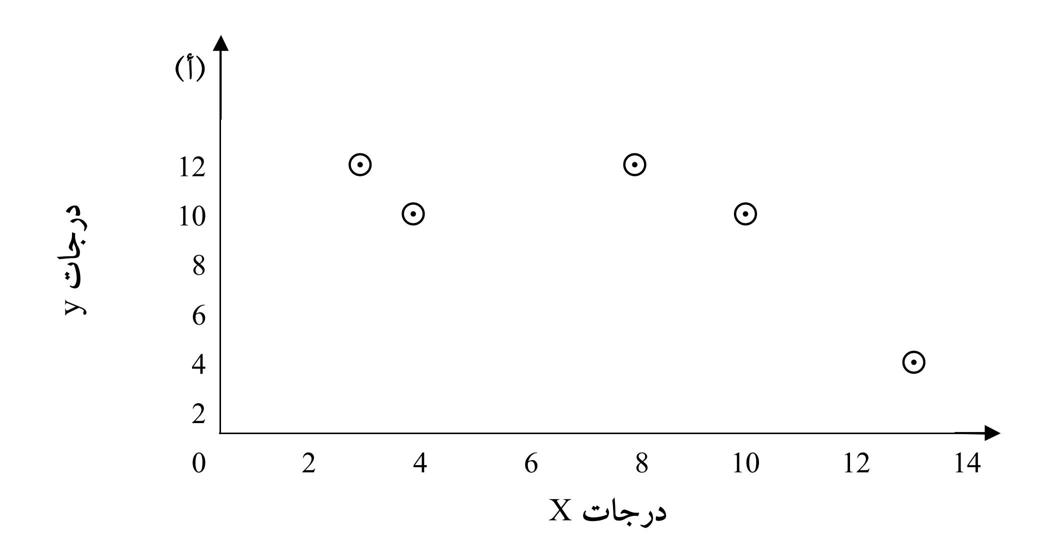
جدول (8-11)

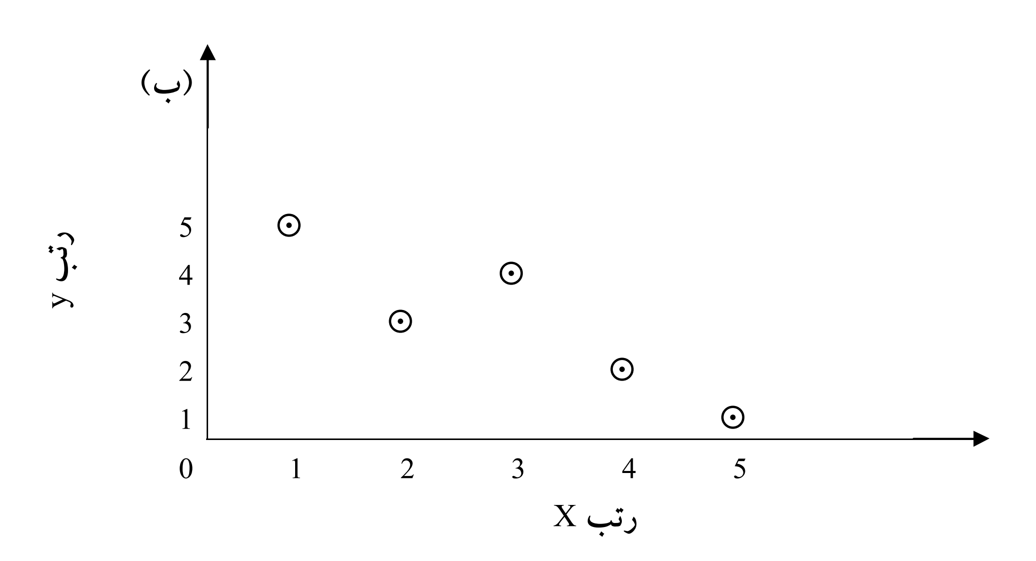
البيانات الأصلية		الرتب		
X	Y	x	Y	xy
3	12	1	5	5
4	10	2	3	6
8	11	3	4	12
10	9	4	2	8
13	3	5	1	5

$$\sum Xy = 36$$

إن شكل الانتشار للبيانات الأصلية والرتب يوضحها الشكل (8 – 5):

ولحساب الارتباط، فإننا نحتاج إلى حساب SS لِـ X و y و Sp. تجدر الإشارة إلى أن كل القيم يتم حسابها مع الرتب، وليس مع البيانات الأصلية. إن رتب X هي ببساطة الأعداد الصحيحة Integers، 1، 2، 3، 4، 5 ويصل مجموع هذه القيم إلى $\sum X = 15$ و $\sum X^2 = 55$.





شكل رقم (8-5) شكل الانتشار والرتب

وأن SS لرتب X هي،

$$SSx = \sum X^2 - \frac{(\Sigma x)^2}{N}$$

$$= 55 - \frac{(15)^2}{5}$$

$$= 55 - 45$$

$$= 10$$

لاحظ أن رتب y مطابقة لرتب X. هذا يعني وجود الأعداد الصحيحة، 1، 2، 3، 4، 5 عليه، فإن SSy ستكون مطابقة لـِ SSx:

$$SSy = \sum y^2 - \frac{(\Sigma y)^2}{N}$$

$$= 55 - \frac{(15)^2}{5}$$

$$= 55 - 45$$

$$= 10$$

ولحساب قيمة SP، فإننا نحتاج إلى: $\sum X$، $\sum y$ و $\sum Xy$ للرتب. إن قيم Xy يحتويها الجدول مع الرتب، كما هو واضح في الجدول (11)، إضافة إلى مجموع كل من X_s و y_s الذي يصل إلى 15. وباستخدامنا لهذه القيم نتـحصل على:

$$Sp = \sum xy - \frac{(\Sigma x)(\Sigma y)}{N}$$

$$= 36 - \frac{(15)(15)}{5}$$

$$= 36 - 45$$

$$= -9$$

وأن الارتباط النهائي لسبيرمان:

$$r_s = \frac{sp}{\sqrt{(ss_x)(ss_y)}}$$

$$= \frac{09}{\sqrt{(10)(10)}}$$

$$= -0.9$$

ويشير ارتباط سبيرمان إلى أن هذه البيانات تبين اتجاه علاقة قوية سالبة (قريبة من العلاقة التامة).

رتب الدرجات المتعادلة Ranking Tied Scores:

عند تحويل الدرجات الأصلية إلى رتب لارتباط سبيرمان، قد يصادف الباحث درجتان أو أكثر متطابقة. وعندما يكون هناك درجتان أو أكثر متطابقة في القيم، فإن الرتب لهذه القيم ينبغي أن تكون متساوية أيضاً. ولإنجاز ذلك ينبغي إتباع الإجراء التالي:

1- تسجل الدرجات في شكل منظم من أصغر الدرجات إلى أكبرها متضمنة القيم المتعادلة في القائمة.

2- تخصيص الترتيب أول، ثانٍ... الخ، لكل وضع في القائمة المنظمة Ordered List.

3- عندما يصادف أن هناك درجتَيْن أو أكثر تكون متعادلة، هنا ينبغي حساب المتوسط لهذه الأوضاع المرتبة. ويخصص قيمة هذا المتوسط كرتبة نهائية لكل درجة.

إن الإجراء المتبع لإيجاد رتب الدرجات المتعادلة يمكن توضيحه من خلال التالي:

الدرجات	الوضع الترتيبي	الرتبة النهائية	
3	1	1.5	المتوسط لـِ 1، 2
3	2	1.5	
5	3	3	
6	4	5	
6	5	5	المتوسط لـِ 4، 5، 6
6	6	5	
12	7	7	

المعادلة الخاصة بارتباط سبيرمان:

بعد أن يقوم الباحث بترتيب القيم المتعلقة بـِ X والقيم المتعلقة بـِ y، فإن عملية حساب SS، و SP عملية ضرورية جداً، أولاً ينبغي على الباحث أن يلاحظ أن رتب X ورتب y هي حقاً مجرد مجموعة من الأعداد الصحيحة: 1، 2، 3، 4،....... n. ولحساب المتوسط لهذه المجموعة من الأعداد الصحيحة يمكننا الحصول على النقطة الوسطى Midpoint لهذه المجموعة من خلال: 2/ (n+1) = M. وبشكل مشابه فإنه يمكننا حساب SS لهذه المجموعة من الأرقام الصحيحة من خلال:

$$SS = \frac{n(n^2 - 1)}{12}$$

أيضاً، وبما أن رتب X ورتب y هي متساوية القيم، فإن SSX سوف تكون متطابقة مع Ssy.

ولما كانت حسابات الرتب بالإمكان تبسيطها، ولما كان ارتباط سبيرمان يستخدم بيانات مرتبة، فإن هذه التبسيطات يمكن دمجها في العملية الحسابية النهائية لارتباط سبيرمان. وبدلاً من استخدام معادلة بيرسون بعد ترتيب البيانات، يمكن للباحث أن يضع الرتب مباشرة في المعادلة البسيطة التالية والتي تعطي نفس النتيجة التي تحصلنا عليها باستخدامنا لمعادلة بيرسون:

$$r_S = 1 - \frac{6\Sigma D^2}{n(n^2 - 1)}$$

حيث إنّ:

D الفرق بين ترتيب X وترتيب y لكل فرد. وينبغي الإشارة هنا إلى أن هذه المعادلة الخاصة يمكن استخدامها فقط بعد أن تحول الدرجات الأصلية إلى رتب. وتستخدم هذه المعادلة فقط عندما لا يكون هناك أي رتب متعادلة. أما إذا كان هناك مجموعة قليلة نسبياً من القيم المتعادلة، يمكن في هذه الحالة استخدام هذه المعادلة، الخاصة، لكن هذه المعادلة ستفقد دقتها كلما زادت الأرقام المتعادلة. ولتوضيح تطبيق هذه المعادلة الخاصة نسوق المثال التالي:

الفرق		الرتب	
D^2	D	Y	X
16	4	5	1
1	1	3	2
1	1	4	3
4	−2	2	4
16	−4	1	5

$$\sum D^2 = 38$$

وباستخدام المعادلة الخاصة لارتباط سبيرمان نتحصل على النتيجة التالية:

$$r_S = 1 - \frac{6\Sigma D^2}{n(n^2 - 1)}$$

$$= 1 - \frac{6(38)}{5(25-1)}$$

$$= 1 - \frac{228}{120}$$

$$= 1 - 1.90$$

$$= -0.90$$

لاحظ أن هذه النتيجة هي بالضبط نفس النتيجة التي توصلنا إليها من خلال استخدامنا لمعادلة بيرسون بالرتب.

ولتفسير سبيرمان راهو، باعتباره مؤشر لقوة التطابق بين المتغيرات، حيث يقع مداه من 0 (لا يوجد تطابق)، إلى ±1 (تطابق تام). وأن قيمة التطابق التام تعني عدم وجود أي اختلافات في الرتب بين المتغيرين (إذا كانت كل الحالات قد رتبت تماماً بشكل منظم ومتساوٍ على كلا المتغيرين).

أما قيمة العلاقة السالبة ($r_S = -1.0$) فهي تعني وجود اختلاف تام. (إذا كانت الحالة رتبت بشكل عالٍ على واحد من المتغيرين كانت أقل على المتغير الآخر... الخ).

إنّ قيمة سبيرمان راهو في هذا المثال وصلت إلى $r_S = -0.90$ تشير إلى علاقة قوية بين هذين، وأن إشارة السالب تشير إلى أن هذه العلاقة علاقة سالبة. وإذا ما تم تربيع هذه القيمة $r_S = -0.90^2$ تساوي $r_S = 0.81$ ، وعليه فإن أخطاء التنبؤ التي تم توليدها، تم تخفيض نسبة 81 %، أي أنه عندما نتنبأ برتبة موضوع ما على المتغير التابع (y)، فإن رتب الموضوع الآخر X تم أخذها في الاعتبار.

الخلاصة:

حاولنا في هذا الفصل أن نبين العمليات التي يتم فيها حساب مقاييس التطابق لنسبة التخفيض من الخطأ (PRE) عندما يكون كلا المتغيرين قد تم قياسهما على الأقل على المستوى الترتيبي. وتجدر الإشارة في هذا السياق إلى أنه لا توجد قاعدة سهلة لتحديد أفضلية أي من هذه المقاييس على الآخر.

إن جزءاً من هذه المشكلة يتعلق بمفهوم التطابق نفسه وإن حقيقة مفهوم التطابق يتم تعريفه إجرائياً بطرق مختلفة فعلى سبيل المثال أن مقياس جاما وتاو هي مقاييس متماثلة في حين أن مقياس سومرز d هو مقياس غير متماثل ولذلك فإن الخيار بالدرجة الأولى يتوقف على النموذج الموجه للعلاقة التي يعتقد الباحث فيها. ولذا فإنه من الناحية العملية فإن هذه المقاييس عادة ما تشير إلى نفس الاتجاه وتقود الباحث للحصول على نتائج متشابهة.

أسئلة للمراجعة:

1- إذا كان هناك نقصان في قيمة متغير يرتبط بزيادة في قيمة المتغير الآخر، فما هو اتجاه هذه العلاقة؟

2- لماذا لا يحدث التطابق بين متغيرين بين كونهما إما تطابق موجب أو تطابق سالب عندما يكون على الأقل واحد من المتغيرات تم قياسه على المستوى الاسمي؟

3- من الجداول التالية: احسب عدد الأزواج المتوافقة. مفترضاً أن الأعداد على حاشية كل جدول تشير للقيم على مقياس ترتيبي:

أ-

3	2	1	
12	24	60	1
8	14	32	2

ب-

3	2	1	
12	24	60	1
8	14	32	2

ج-

4	3	2	1	
42	25	17	12	1
24	19	14	10	2
20	16	11	6	3

د-

4	3	2	1	
42	25	17	12	1
24	19	14	10	2
20	16	11	6	3
22	14	9	3	4

4- من الجدول التالي احسب عدد الأزواج غير المتوافقة مفترضاً أن الأعداد في الحاشية تشير إلى قيم مقياس ترتيبي:

أ-

	1	2	3
1	60	24	12
2	32	14	8

ب-

	1	2	3
1	60	24	12
2	32	14	8

ج-

	1	2	3	4
1	12	17	25	42
2	10	14	19	24
3	6	11	16	20

5- من البيانات التالية التي تبين العلاقة بين التدخين ومستوى الصحة لدى مجموعة من الناس:

مستوى الصحة العامة	**عادة التدخين**		
	لا يدخن	**يدخن**	**المجموع**
سيئة	13	34	47
جيدة بعض الشيء	22	19	41
جيدة	35	9	44
جيدة جداً	27	3	30
المجموع	**97**	**65**	**162**

المطلوب:

أ- بالنظر إلى توزيع الصفوف هل بالإمكان استبيان التطابق بين هذين المتغيرين؟ وما هو اتجاه هذا التطابق؟ هل هذا الاتجاه يظهر عند حساب مقياس التطابق؟

ب-احسب قيمة جاما Gamma، وسومرز Somerś d d وبين ما النتيجة التي توصلت إليها حول العلاقة بين التدخين ومستوى الصحة العامة لدى هؤلاء الناس؟

6- من البيانات التالية احسب قيمة جاما وسومرز d. وفسر هذه النتائج التي تتوصل إليها:

غياب الأم من البيت	مستوى الإنجاز لدى الطفل			المجموع
	ضعيف	جيد	ممتاز	
باقية في البيت (لا تعمل)	20	58	22	100
تعمل جزئياً	15	62	23	100
تعمل طوال النهار	12	62	26	100
المجموع	**47**	**182**	**71**	**300**

7- من البيانات التالية استخدم سبيرمان الخاصة لحساب الارتباط. وأن هذه الدرجات D^2 يساوي 0، 4، 0، 1، 1. وأن مجموع $\sum D^2 = 6$.

الدرجة X	الدرجة y	الرتبة X	الرتبة y	D
5	12	2	2	0
7	18	3	5	2
2	9	1	1	0
15	14	5	4	−1
10	13	4	3	−1

8- من البيانات التالية احسب ارتباط بيرسون مبيناً الخطوات المتبعة؟

	Yx	Y	x	الأشخاص
SSX=40	0	0	0	1
SSy=54	2	1	2	2
	80	10	8	3
	54	9	6	4
	24	6	4	5
	160	30	20	

الهوامش والمصادر:

أولا: الهوامش:

1- George Argyrous, Statistics for Social and Health Research With a Guide to SPSS, Sage Publications ,London, 2001, PP. 173 - 176.

2- Ibid, PP. 179 - 178.

3- Ibid, PP. 180 - 182.

4- Frederick J. Gravetter and Larry B. Wallnau, Statistics for The Behavioral Sciencies, 8th ed, Wadsworth Cengage Learning, USA, 2010, PP. 540 - 547.

ثانيا: المصادر:

1- Frederick J. Gravetter and Larry B. Wallnau, Statistics for The Behavioral Sciencies, 8th ed, Wadsworth Cengage Learning, USA, 2010.

2- George Argyrous, Statistics for Social and Health Research With Guide to SPSS, Sage Publications, London, 2001 .

3- Joseph F. Healey, The Essentials of Statistitics: A Tool for Social Research, Wadsworth Cengate Learning, USA. 2010.

الفصل التاسع

التطابق بين متغيرات تم قياسها على المستوى ذي المسافات والنسبي: شكل الانتشار وخط الانحدار

مقدمة

تناولنا في الفصلين السابقين طرق وصف البيانات عندما يكون أحد المتغيرات على الأقل قد تم قياسه على المستوى الاسمي أو الترتيبي، وقد تم وصف هذه البيانات في شكل جداول تقاطع. ومن خلال جداول التقاطع يتبين لنا التوزيع المشترك لمتغيرين نستطيع من خلالهما ملاحظة ما إذا كان هناك تطابق بين المتغيرين. وعند فحص التوزيعات التكرارية النسبية في الجدول، يقودنا إلى أن نشك في أنَّ هذين المتغيرين يرتبط بعضهما بالبعض الآخر، مما يحدو بنا كخطوة لاحقة بحساب مقاييس التطابق التي تمكننا من الحصول على قيمة رقمية محددة لمثل هذا الشك. وتجدر الإشارة إلى أنه إذا كانت البيانات لهذين المتغيرين تحت الدراسة قد تم جمعها على المستوى ذي المسافات والنسبي، الذي يعني عادة وجود عدد كبير من القيم، الأمر الذي يصبح فيه التعامل مع جداول التوافق غير مناسب لوصف التوزيع. ومن هنا فإن التقنية المرادفة لجداول التوافق للبيانات المناسبة المقاسة على المستوى ذي المسافات / النسبي هي الرسم الانتشاري.

الرسم الانتشاري Scatter Plots:

إنه من الصعوبة بمكان أن نرتب البيانات ذات المسافات المتساوية والنسبية في جداول تقاطع؛ فالبيانات ذات المسافات المتساوية والنسبية عادة لا تقع في عدد صغير من الفئات المنفصلة صغير السن، كبير السن... الخ كالبيانات ذات حجم صغير أو كبير.

إن مثل هذه البيانات بطبيعة الحال يمكن اختزالها إلى قيم قليلة، ولكن هذا الاختزال سيكون على حساب المعلومات. وحيث إن هناك عادةً مجموعة من القيم للمتغيرات التي تم قياسها على المستوى ذي المسافات والنسبي، فإن جدول التوافق سوف يحتوي على عدد كبير من الصفوف والأعمدة، كما توجد عدة قيم في البيانات. وإذا نظرنا إلى التوزيع العمري بالسنوات لسكان بلد ما فإننا نحتاج إلى أكثر من 100 صف من البيانات لكي نأخذها في الاعتبار، فالحقيقة القائلة إن العمر سوف يتوزع على مدى واسع. إن رسم البيانات في شكل انتشاري يمكننا من الحصول على مدى واسع من القيم التي توفر لنا عادة أفضل طريقة لتنظيم مثل هذه البيانات للحصول على انطباع أولي فيما إذا كان هناك وجود لأي علاقة.

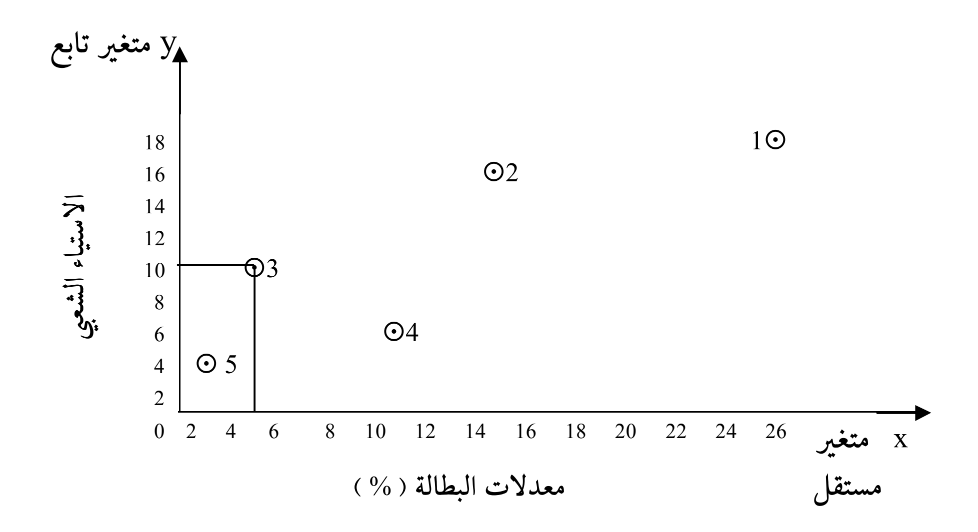
إن رسم الانتشار يبين لنا مجموعة من القيم المتآلفة بحيث إنّ كل حالة تسجل على كلا المتغيرين بشكل تلقائي. إن رسم الانتشار يبين لنا التوزيع المشترك لمتغيرين متصلين ومتناسقين على رسم الانتشار الذي يشير إلى القيم لكل حالة تدون لكل واحد من المتغيرين. فعلى سبيل المثال، قد نرغب في معرفة العلاقة بين معدلات البطالة ومستوى الاستياء الشعبي عن المدن والتي يمكن الحصول عليها من الإحصاءات الرسمية المتعلقة بمعدلات البطالة. فالبطالة لمتغير مستقل X وعدد حالات الاستياء الشعبي لمتغير تابع y لخمس مدن.

جدول (9 ـ 1) معدلات البطالة والاستياء الشعبي في خمس مدن

المدن	معدلات البطالة	معدلات الاستياء
1	25	17
2	13	15
3	5	10
4	10	5
5	2	4

المصدر : George Argyrous,, Statistics for Social and Health Research, With a Guide to Spss, op.cit , p. 233

وبترتيب هذه المعلومات في رسم انتشاري شكل (9 – 1) تجعل من السهولة بمكان قراءتها لتحديد فيما إذا كانت هناك علاقة.



شكل رقم (9 ـ 1) الرسم الانتشاري لبيانات البطالة والاستياء الشعبي

لقد جرت العادة أن يوضع المتغير التابع (y) على المحور الرأسي والمتغير المستقل (X) على المحور الأفقي عند بناء شكل الانتشار. وإذا ما نظرنا إلى أي واحدة من هذه النقاط 1-5 ورسمنا خط مستقيم من الأسفل إلى المحور الأفقي فإنه بإمكاننا أن نجد معدل البطالة في تلك المدينة. وبشكل مشابه برسمنا لخط مستقيم عبر المحور الرأسي يمكننا قراءة عدد حالات الاستياء الشعبي، فالخط المتقاطع لمدينة 3 كما تم توضيحه في هذا الإجراء يشير إلى أن معدلات البطالة لهذه المدينة هو 5 % وأن عدد حالات الاستياء الشعبي هي أيضاً 10 أحداث من الاستياء الشعبي. وبالنظر إلى هذا الشكل فإننا بداهة نرى أن هناك تطابق، لأننا استطعنا أن نتصور ميل الخط ماراً خلال هذه النقاط الخمس. وأن اتجاه التطابق يشار إليه من خلال ما إذا كانت ميول الخط التخيلي أعلى (موجب) أو أسفل (سالب). في هذه الحالة التي أمامنا فإن ميل الخط يكون موجباً مشيراً إلى أن الزيادة في معدلات البطالة متطابقة مع الزيادة في عدد حالات الاستياء الشعبي[1].

خط الانحدار Regression Line:

يعني الانحدار ببساطة مطابقة الخط خلال الرسم الانتشاري للحالات التي تكون مطابقة بشكل جيد للبيانات. وأن أي خط يمكن توضيحه من خلال المعادلة الرياضية التالية المتعلقة بالخط المستقيم. حيث إن: $y = a \pm bX$.

حيث أن: y المتغير التابع.

و X المتغير المستقل.

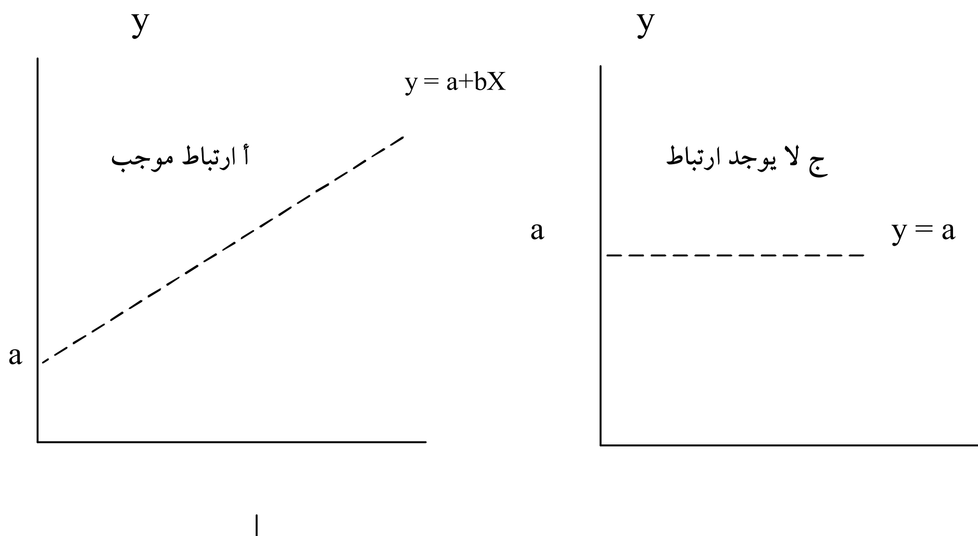
و a قيمة ثابتة وهي المساقة على المحور y من البداية إلى النقطة التي تقطع محور y فهي قيمة y عندما تكون X تساوي 0.

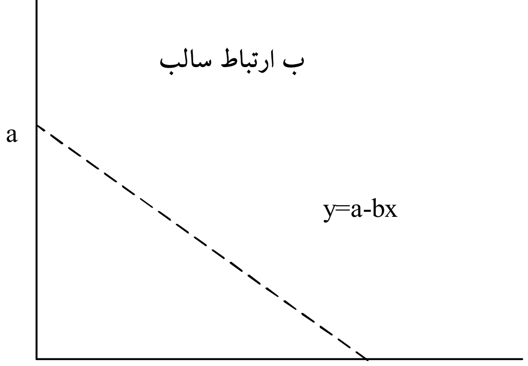
B تشير إلى ميل الخط.

+ تشير إلى التطابق الموجب.

– تشير إلى التطابق السالب.

من خلال هذه المعادلة يمكننا القول بأنَّ الخط المستقيم يمكن تعريفه من خلال عاملين أساسيين أولهما: بداية بنقطة على طول المحور الرأسي، a، وثانيهما هو ميل الخط من نفس هذه النقطة b ± وأن قيمة b هي القيمة التي نرغب فيما إذا كان أي ميل إما أن يكون موجباً أو يكون سالباً، كما تشير لبعض الارتباط بين المتغيرين. ففي الشكل (9 – 2) فإننا نلاحظ ثلاثة خطوط مختلفة تعكس قيمة b في ثلاثة مواقف بديلة لارتباط موجب، وارتباط سالب، ولا يوجد ارتباط.



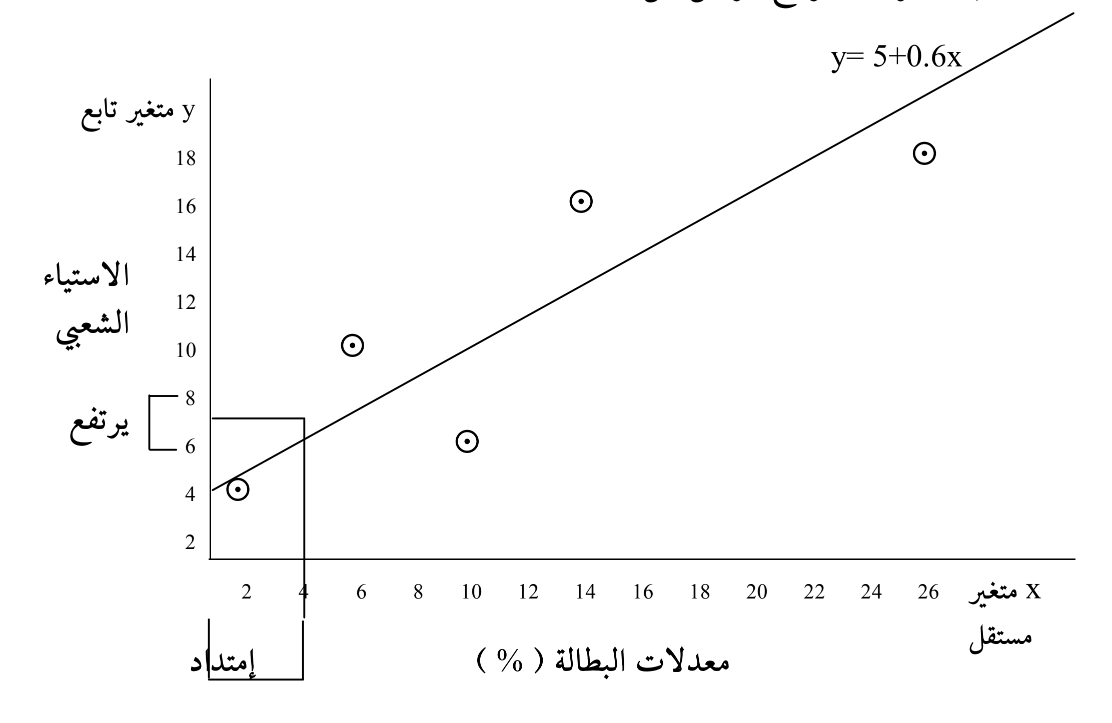


شكل رقم (9-2): ثلاث خطوات تبين ارتباط موجب أ وارتباط سالب ب ، (ج) لا يوجد ارتباط 0

وبالنظر إلى البيانات المتعلقة بالمدن الخمس يمكننا رسم مجموعة من الخطوط المستقيمة من خلال الشكل الانتشاري، وأن كل واحد من هذه الخطوط سوف يحتوي على معادلته الفريدة والخاصة به. فعلى سبيل المثال، في شكل (9 – 2) قد تم رسم خط يبدو لنا أنه يتوافق مع البيانات بشكل جيد. ويمكن أن نطلق على هذا الخط خط 1 وبدلاً من ذلك يمكننا التعبير عنه بمعادلة رياضية:

$$y = 5 + 0.6X$$

إن السؤال المطروح هو من أين جاءت هذه المعادلة؟



شكل رقم (9-3)

القيمة لِـ: a (5) وهي النقطة على محور y عندما يبدأ الخط. وهي عدد حالات الاستياء الشعبي الذي نتوقع أن نجده في مدينة لديها معدل بطالة صفر (0).

- علامة + تعني أن الخط لديه ميل موجب الذي يشير إلى ارتباط موجب بين هذين المتغيرين.
- أما قيمة 0.6 لـِ b فهي تشير إلى الميل، أو معامل خط الانحدار.

إن معامل خط الانحدار تشير إلى كم من الاستياء الشعبي سيزيد إذا ما زادت معدلات البطالة بـِ 1 % ولما كان الميل لأي خط مستقيم هو الارتفاع على الخط الممتد، فإنه بإمكاننا حساب قيمة (1) لأي ارتفاع في الاستياء الشعبي مثل زيادة 3 بين 5-8. حينئذٍ يمكننا قراءة الزيادة المقابلة في معدلات البطالة، التي تعطينا امتداد خمس (5) وبقسمة الارتفاع على الامتداد (run) فإن الميل سيكون كالتالي:

$$b = \frac{\text{الارتفاع} rise}{\text{الامتداد} run} = \frac{3}{5}$$

$$= 6.$$

ن الخط الذي أوضحناه الآن يعطينا مدى القيم المتوقعة للاستياء الشعبي اعتماداً على قيمة معدلات البطالة وإن الفرق بين القيمة المتوقعة والقيمة الفعلية للاستياء الشعبي حول معدل بطالة محدد، يطلق عليه المتبقي residual أو مصطلح الخطأ Error Term. إن المتبقي أو مصطلح الخطأ هو الفرق بين القيمة التي نتحصل عليها من خلال المشاهدة للمتغير التابع وبين القيمة المتوقعة للمتغير نفسه بواسطة أو عبر خط الانحدار. لاحظ أن الخط المستقيم لن يمر خلال كل النقاط في الشكل الانتشاري. في حقيقة الأمر أن الخط الجيد من الممكن أنه لا يمس أي من هذه النقاط: عادةً ما تُوجد فجوة بين كل رسم بياني وخط الانحدار باستثناء النقطة التي تقع على الخط ومن هنا تكون لدينا قيمة متبقية. على سبيل المثال، أن الخط الذي لدينا يتكهن بأن المدينة رقم 4 بمعدل بطالة 10 %، وعدد الاستياء الشعبي في هذه المدينة سيكون كالتالي:

$$y = 5 + 0.6X = 5 + 0.6(10)$$

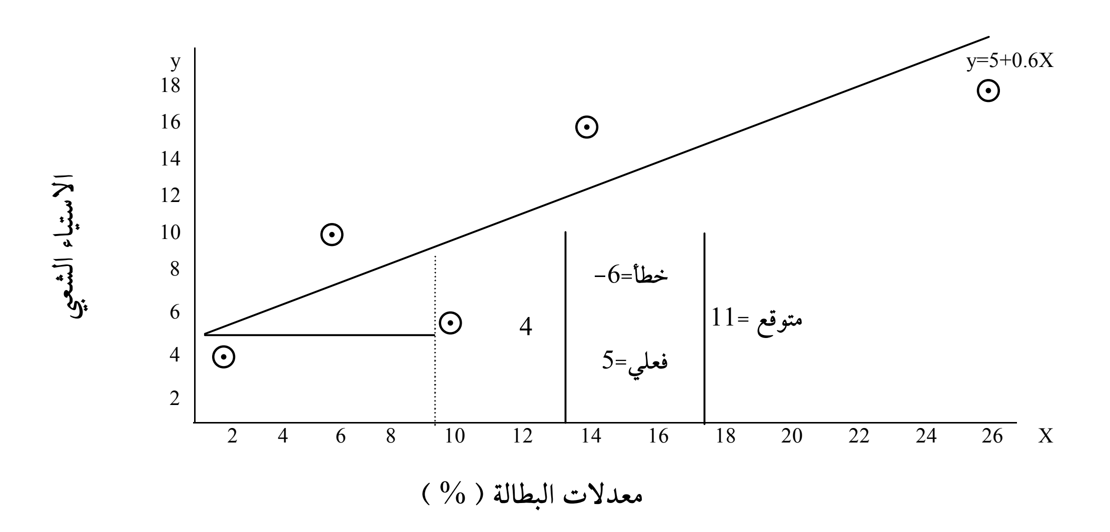
$$= 11$$

وبدلاً من ذلك، يوجد لدينا خمس حالات من الاستياء الشعبي لمدينة 4 بمعدل بطالة يصل إلى 10 %. وبالتالي فإن الخطأ (e) عند هذه النقطة يصل إلى 6-:

e = y - y = 5 - 11 المتوقعة الفعلية

= -6

والشكل التالي يوضح هذه المسألة:



شكل (9-4) الدرجات الفعلية والدرجات المتوقعة

واستناداً على الخط الذي تم رسمه فإننا نلاحظ من خلال العين المجردة أن الخط إلى حد بعيد يطابق هذه البيانات. إن أفضل خط هو ذلك الخط الذي يقلل من البواقي ويجعلها قليلة قدر الإمكان. أي الخط الذي يقلل إلى حد أدنى البواقي.

إن تحليل الانحدار يستخدم الفكرة الرئيسية وهي: أقل مربعات الانحدار Least Square Regression (LSR) لتقليص الفارق بين القيمة المتوقعة لِـ y والقيم الفعلية لِـ y (مربعة) على أن يكون هذا الفرق صغيراً قدر الإمكان (عادة ما نربع البواقي أكثر من أن نقوم بجمعها حيث إن مجموع البواقي لأي خط تمر خلاله النقطة التي يكون فيها المتوسط

لكلا المتغيرين التابع والمستقل مساوياً لصفر. ولتقليل أثر العلامة الموجبة والسالبة فإنه يتم تربيع البواقي بحيث يصبح تعاملنا مع الأعداد الموجبة).

إن أقل مربعات الانحدار هي تلك القاعدة التي تبين لنا رسم الخط خلال الرسم الانتشاري والذي يقلل مجموع البقايا المربعة. إنه بإمكاننا إيجاد أقل مربعات خط الانحدار خلال عملية التجربة والخطأ فإنه بإمكاننا أن نستمر في رسم الخطوط خلال شكل الانتشار مع إيجاد المعادلات الخاصة والبواقي حتى نصل في نهاية الأمر إلى الخط الذي يقلل من هذه البواقي. ولحسن الحظ يوجد لدينا بديل عندما نستخدم القاعدتين الاثنتين حتى نستطيع اشتقاق أقل مربعات خط الانحدار بشكل مباشر. إن خط المربعات الصغرى للانحدار ينبغي أن تمر خلال النقطة التي تتساوي فيها المتوسطات للمتغير التابع والمتغير المستقل ($\bar{x}$ و $\bar{y}$). والمتوسط لعدد حالات البطالة $\bar{x}$ تكون:

$$\bar{x} = \frac{\sum x}{N} = \frac{55}{5}$$

$$= 11$$

وأن متوسط معدل الاستياء الشعبي $\bar{y}$

$$\bar{y} = \frac{\sum y}{N} = \frac{55}{5}$$

$$= 10.2$$

وعليه فإن أقل مربعات الانحدار خط الانحدار سوف تمر خلال النقاط المتناسقة (11 و 10.2) ويمكننا تعريف ميل خط الانحدار b من خلال المعادلة التالية:

$$b = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\Sigma(x - \bar{x})^2}$$

وحيث إنَّ هذه المعادلة تبين الفكرة الأساسية التي يحتاجها الخط للتقليل من فروق التربيع بين القيم الفعلية والقيم المتوقعة، من هنا فإنه من السهولة بمكان حساب قيمة b وفقاً للمعادلة التالية:

$$b = \frac{N\,\Sigma(xy) - (\Sigma x)(\Sigma y)}{N\,\Sigma x^2 - (\Sigma x)^2}$$

جدول (9ـ2) حساب ميل خط الانحدار لأقل المربعات

المدن	معدلات البطالة X	حالات الاستياء y	X^2	y^2	Xy
1	25	17	625	289	425
2	13	15	69	225	195
3	5	10	25	100	50
4	10	5	100	25	50
5	2	4	4	16	8
	$\sum X55$	$\sum y = 51$	$\sum x^2 = 923$	$\sum y^2 = 655$	$\sum Xy = 728$

$$b = \frac{N\sum(xy) - \sum(x)\sum(y)}{N\sum x^2 - (\sum x)^2} = \frac{5(728) - (55)(51)}{5(923)(55)^2}$$

$$= \frac{3640 - 2805}{4615 - 3025}$$

$$= +\,0.53$$

إن قيمة b يطلق عليها معامل الانحدار، وهي قيمة ذات أهمية لأنها تحدد مقدار أي ارتباط بين متغيرين. فمعامل الانحدار يشير إلى كم من الوحدات في المتغير التابع ستتغير إذا ما تغيرت أي وحدة في المتغير المستقل.

وتجدر الإشارة هنا إلى أنه بعدما حددنا خط الانحدار خلال نقطة محددة (متوسط X و y) وكذلك من خلال إعطاء الاسم الأخير من خلال حساب ميل الخط خلال هذه النقطة، فإنه باستطاعتنا إعطاؤها وصفاً كاملاً من خلال اشتقاق القيمة المتعلقة بـِ a. ومن هنا يمكننا استخدام المعادلة التالية التي تستخدم كلا المَعْلَمَيْن المتعلقين بخط الانحدار الذي

تم بيانه (تلك التي تمر خلال المتوسط المتعلق بـِ X والمتوسط المتعلق بـِ y وبالتالي يكون لها ميل مساوياً لـِ b:

$$a = \overline{y} - b\overline{x}$$

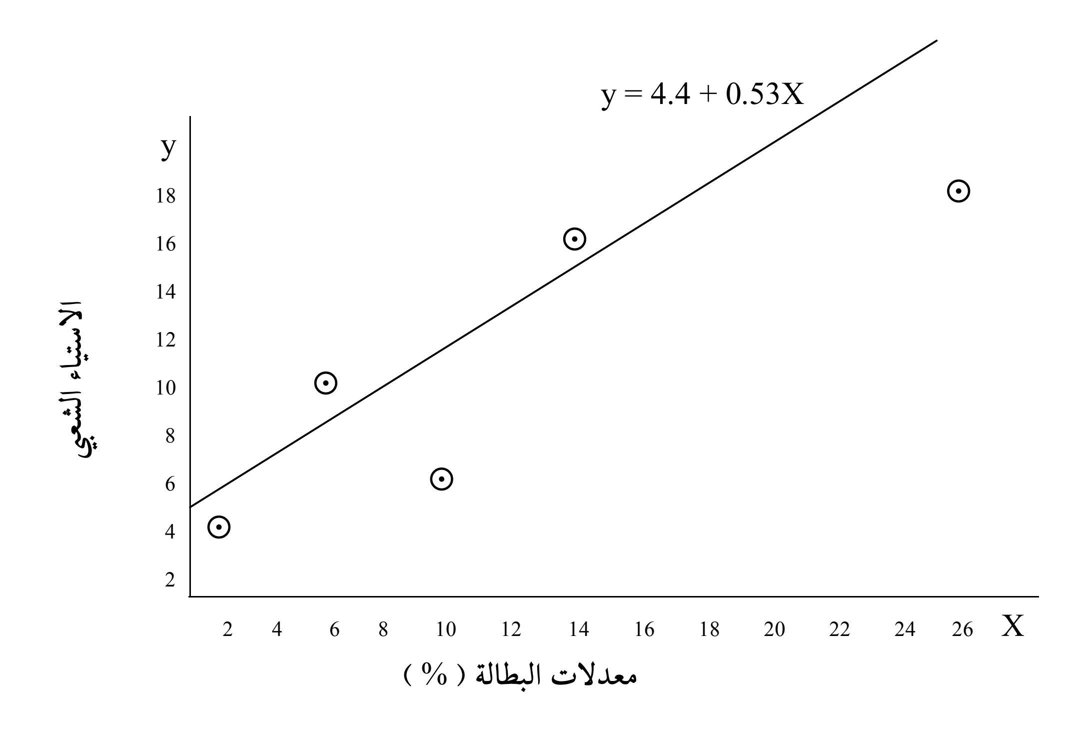
وعليه فقيمة a ستكون:

$$a = \overline{y} - b\overline{x} = 10.2 - 0.53(11)$$

$$= 4.4$$

وعليه يمكننا تعريف أفضل خط مقابل لمثل هذه الحالات من خلال المعادلة التالية:

$$y = 4.4 + 0.53X$$



شكل (9-5) خط الانحدار لأقل المربعات (OLS)

السؤال الذي يمكن طرحه هنا هو: ماذا يخبرنا خط الانحدار لأقل المربعات حول العلاقة بين معدلات البطالة وحالات الاستياء الشعبي لمثل هذه المجموعة من الحالات؟ للإجابة على هذا السؤال يمكننا القول:

- توجد علاقة موجبة بين المتغيرين: فالزيادة أو النقصان في معدل البطالة يرتبط مع الزيادة أو النقصان في عدد حالات الاستياء الشعبي.
- يمكننا قياس العلاقة الموجبة: فالزيادة في معدل البطالة لـِ: 1% يرتبط بالزيادة بـِ 0.53 لحالات الاستياء الشعبي.

إنه بإمكاننا الآن استخدام هذه المعادلة لغرض التنبؤ بعدد حالات الاستياء الشعبي في مدينة ما، فمن المرجح أن يكون لديها معدل محدد من البطالة، على سبيل المثال إذا ما علمنا أن مدينة أخرى لديها معدل بطالة يصل إلى 18 % فإن أفضل تكهن يمكن الوصول إليه هو أنها تمر بتجربة 13.9 حالة استياء شعبي [2].

معامل ارتباط بيرسون (r) :

قد تبين لنا من خلال المثال السابق أن قيمة b هي مؤشر ما إذا كانت هناك علاقة بين متغيرين تم قياسهما على مستوى المقياس ذي المسافات والنسبي، وكذلك تبين اتجاه مثل هذا الارتباط. ولكن السؤال هل تشير أيضاً هذه القيمة إلى قوة العلاقة؟ هل قيمة b التي تساوي 0.53 تشير إلى تطابق، قوي أو متوسط أو ضعيف؟ إنَّ هذه القيمة لسوء الحظ لا تشير إلى ذلك.

إن القضية الأساسية هي أن وحدات القياس تتباين من موقف إلى موقف آخر. فعلى سبيل المثال، إذا استخدمنا التناسب بدلاً من نقاط النسب لقياس معدلات البطالة، فعوضاً عن استخدام الحسابات 22، 20، 15، 10، 9 يمكننا استخدام 0.22، 0.20، 0.15، 0.1، 0.09 فإن القيمة المتوقعة لـِ: b ستكون 53 بدلاً من 0.53.

إن العلاقة الفعلية التي نود معرفتها لم تتغير، وإنما تغيرت فقط وحدات القياس. بمعنى آخر، إن قيمة b قد تأثرت ليس فقط بقوة الارتباط، وإنما أيضاً بوحدات القياس. وعليه لا توجد طريقة لمعرفة ما إذا كانت أي قيمة محددة لِـ: b تشير إلى ارتباط ضعيف أو متوسط أو قوي. وللتغلب على هذه المشكلة يمكننا تحويل قيمة b إلى مقياس معياري للارتباط يطلق عليه معامل ارتباط بيرسون (r). إن معامل بيرسون (r) يتراوح مداه دائماً بين −1 و +1 بغض النظر عن الوحدات الفعلية التي تم قياس المتغيرات عليها. والمعادلة المستخدمة لارتباط بيرسون (r) هي:

المعادلة (1)

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{[(x-\bar{x})^2][(y-\bar{y})^2]}}$$

أو

المعادلة (2)

$$r = \frac{N\Sigma(xy)-(\Sigma x)(\Sigma y)}{\sqrt{[N\Sigma x^2-(\Sigma x)^2][N\Sigma y^2-(\Sigma y)^2]}}$$

وإذا نظرنا إلى المثال الذي تم التعامل مع مكوناته من خلال حساب قيمة b جدول (2) فإذا ما استبدلنا الإحصاءات من هذا الجدول وتعويضها بالمعادلة التالية سنتحصل على:

$$r = \frac{N\Sigma(yx)-(\Sigma x)(\Sigma y)}{\sqrt{[N\Sigma x^2-(\Sigma x)^2][N\Sigma y^2-(\Sigma y)^2]}}$$

$$= \frac{5(728-55(51)}{\sqrt{[5(923)-(55)^2][5(655)-(51)^2]}}$$

$$= \frac{3640-2805}{[4615-3025][3275-2601]}$$

$$= 0.81$$

إما إذا استخدمنا المعادلة الأولى لإيجاد ارتباط بيرسون (r) فإننا نتحصل على نفس القيمة كما هو موضح في التالي:

1	2	3	4	5	6	7
X المدن	y البطالة	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
25	17	+14	+6.8	95.2	196	46.24
13	15	+2	+4.8	9.6	4	23.04
5	10	-6	-0.2	1.2	36	0.04
10	5	-1	-5.2	5.2	1	27.04
2	4	-9	-6.2	55.8	81	38.44
$\Sigma x = 55$	$\Sigma y = 51$	$\sum(x-\bar{x}) = 00$	$\sum(y-\bar{y}) = 00$	167	$\sum(x-\bar{x})^2 = 318$	$\sum(y-\bar{y})^2 = 138.8$

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{[(x-\bar{x})^2][(y-\bar{y})^2]}}$$

$$= \frac{167}{\frac{\sqrt{(318)(134.8)}}{\sqrt{42866.4}}}$$

$$= \frac{167}{207.42}$$

$$= 0.81$$

إن قيمة r تبين لنا فقط قوة التطابق بل أيضاً اتجاه هذا التطابق. إن قيمة 0.81 تشير إلى الارتباط بين هذين المتغيرين لمثل هذه المجموعة من الحالات فهي قيمة موجبة وقوية.

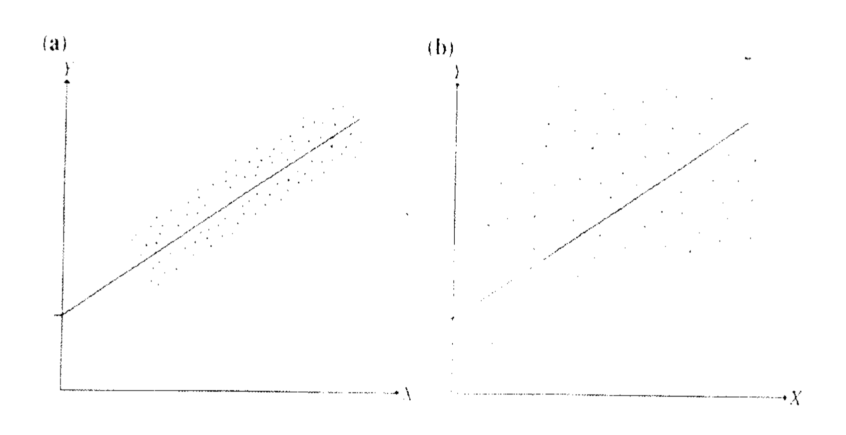
تفسير التباين: معامل التحديد أو التقدير:

قد بينا في هذا الفصل في حينه استخدام خط الانحدار للتكهن بعدد حالات الاستياء الشعبي في مدينة ما من خلال معطى معين لمعدل البطالة. إلا أننا قد رأينا أن هناك دائماً

"هامش للخطأ" في هذا التنبؤ، اعتماداً على مدى قرب النقاط وتجمعها حول خط الانحدار. من هنا يمكننا استخدام خط الانحدار لنبين أن زيادة معينة في X ستقود إلى زيادة كبيرة في y، ولكن إذا كان هناك خطأ كبير بين خط الانحدار والنقاط الفعلية للبيانات فالاحتمال القوي أن التنبؤات سوف تكون خاطئة بقدر كبير أكثر إذا قورنت بالموقف عندما تكون الدرجات متجمعة بشكل واضح حول خط الانحدار.

وكما يظهر في الشكل التالي أنه بالرغم من أن نفس خط الانحدار يتطابق بشكل واضح لكلتا المجموعتين من النقاط فإن ثقتنا ستكون كبيرة في قدرتنا للتنبؤ بقيمة (a) أكثر من قدرتنا على التنبؤ بقيمة (b) ويرجع السبب في ذلك إلى أن خط الانحدار في (a) قادر على تفسير قدر كبير في نسبة التباين في (y) أكثر منه في خط الانحدار (b).

وعليه فإننا بحاجة إلى بعض القياس لكي يبين لنا كم من التباين في المتغير التابع يتم تفسيره بواسطة خط الانحدار.



شكل (9-6)

إنه من حسن الحظ أنه يمكننا أن نقوم بذلك ببساطة من خلال تربيع قيمة (r) للحصول على قيمة معامل التحديد r^2، فالتباين المفسر من خلال خط الانحدار ذا صلة بالتباين المفسر في الحالة التي يوجد بها تطابق.

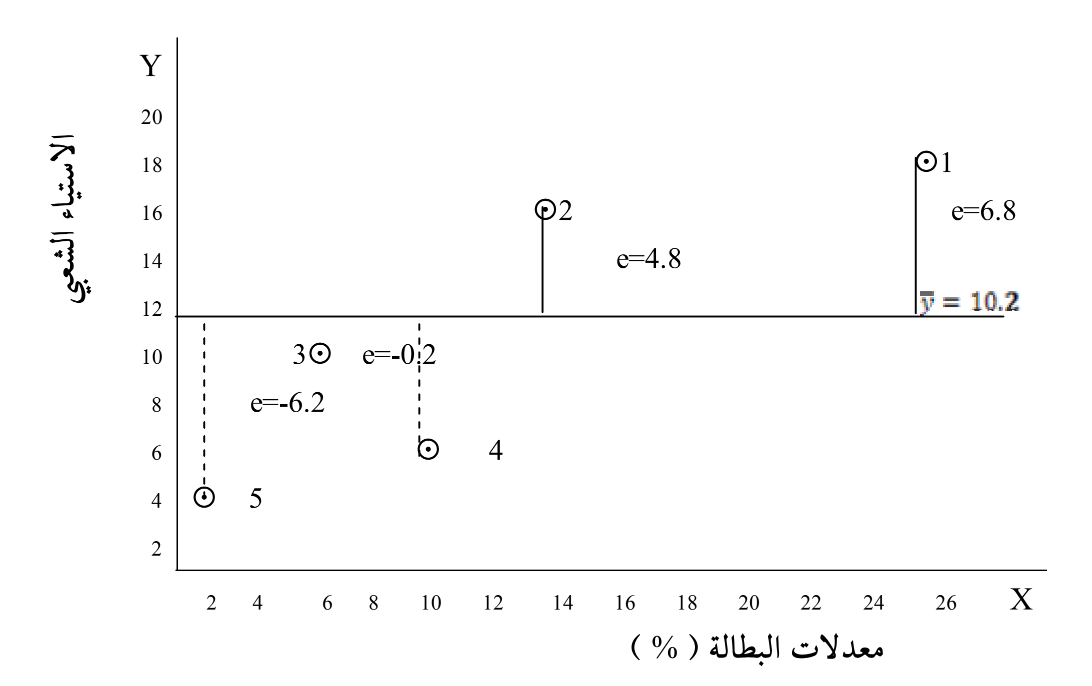
$$r^2 = (0.8!)^2$$
$$= 0.65$$

ويمكننا تفسير معامل التحديد كمقياس غير متماثل للتقليل من نسبة الخطأ في مقياس التطابق أي أن معامل التحديد تشبه إلى حد كبير مقاييس نسبة التقليل من الخطأ Proportional reduction in Error (PRE) التي تم التعرض لها في موضعه في الفصلين السابقين وبالتالي فإن معامل التحديد لديه منطق مشابه لمنطق لامبيدا، إلا أن تطبيقات معامل التحديد تتعلق بالمتغيرات المقاسة على المستويين ذي المسافات والنسبي.

ومن هنا يمكننا القول بأن معامل التحديد يُستخدم لتحديد الدلالة على مقدار مؤاءمة خط أقل التربيعات مع بيانات العينة الذي يسعى الباحث لاختبارها. وتُعرف معامل التحديد بأنه "النسبة بين مجموع مربعات الانحراف الراجح إلى الانحدار إلى مجموع مربعات الانحراف الكلية"[3]، ومن هنا نجد أن معامل التحديد أو التقدير يبين لنا نسبة الاختلاف الكلية في المتغير التابع (y) والتي تم تفسيرها من انحدار y على X.

وتجدر الإشارة إلى أنه أيضاً باستطاعتنا القيام بالتنبؤ حول القيم المتوقعة للمتغير التابع بدون أية معلومات حول المتغير المستقل (X). كما أننا إذاً يمكننا القيام بعملية التنبؤ عندما تكون لدينا معرفة حول المتغير المستقل (X)، ومقارنة معادلات الخطأ فعلى سبيل المثال، إذا أردنا أن نخمن في عدد حالات الاستياء الشعبي في كل مدينة وأن كل ما نعرفه أن متوسط عدد حالات الاستياء الشعبي لكل المدن الخمس هو 10.2، فإن أفضل توقع يمكن الوصول إليه هو أن عدد حالات الاستياء الشعبي لكل مدينة هو 10.2 بغض النظر عن المعدل الفعلي للبطالة [4].

نقوم برسم خط أفقي مستقيم لهذه القيمة كخط انحدار خلال شكل الانتشار. انظر الشكل التالي:



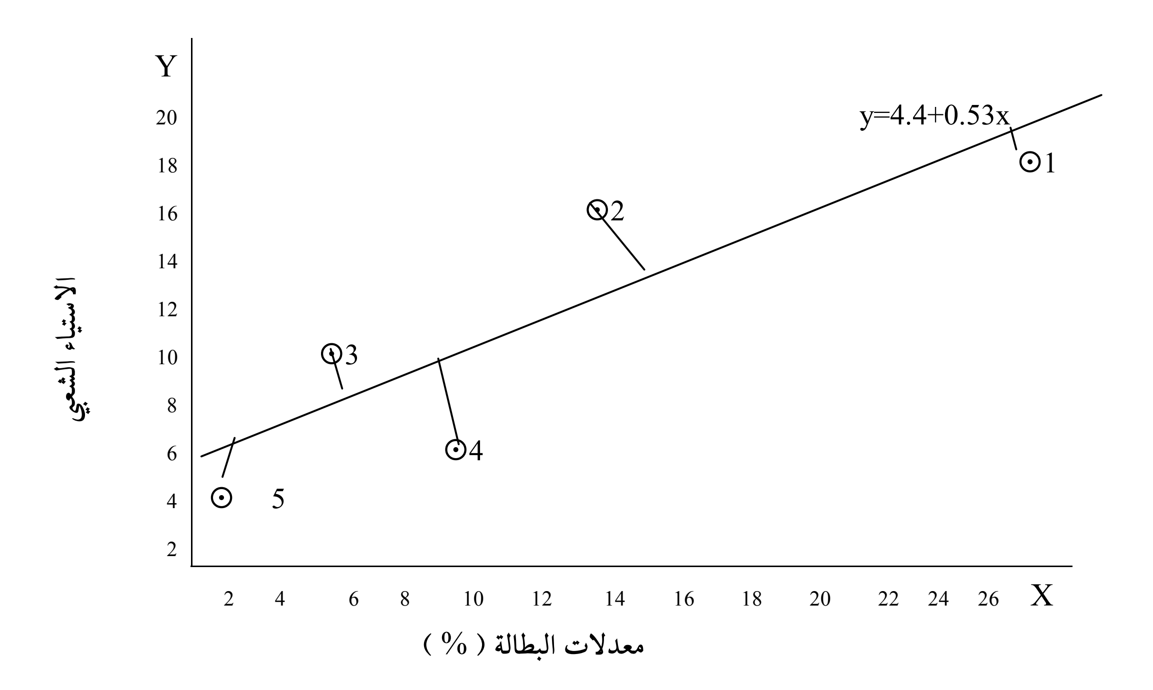
شكل (9-7) خط الانحدار بدون معلومات حول المتغير المستقل

إن هذا الخط الأفقي هو الخط الذي تم رسمه عندما لم يكن هناك ارتباط بين هذين المتغيرين (معدل البطالة وعدد حالات الاستياء الشعبي). إن معرفة ما إذا كان معدل البطالة عالياً أو منخفضاً لن يقودنا إلى تغيير توقعاتنا حول حالات الاستياء الشعبي عن المتوسط.

ففي بعض الأحيان قد يأتي هذا الخط بشكل قريب من العلاقة. ففي مدينة 3 يمكننا أن نرى أن هذا الخط يتنبأ بمعدل بطالة يصل إلى 5 %، والتي ستكون 10.2 من حالات الاستياء الشعبي.

فهنا نجد أنه في الحقيقة أن 10 حالات من الاستياء الشعبي. تولد لنا خطأ (e) في هذه المدينة يصل إلى 0.2 فقط. وعلى أية حال، فإنه في شواهد أخرى يمكننا أن نسبب خطأ أكبر باستخدام هذا الخط. ففي مدينة 1 على سبيل المثال، يصل معدل البطالة فيها

إلى 25 % مرة ثانية فإننا نتنبأ بأنه في هذه المدينة تصل حالة الاستياء الشعبي فيها إلى 10.2 لكنه في الواقع نجد أن حالات الاستياء الشعبي في هذه المدينة يصل إلى 17 حالة استياء شعبي. ومن هنا نجد أننا قد ارتكبنا خطأً في التكهن بمعدلات الاستياء الشعبي يصل إلى 6.8. والآن إذا قارنا هذه الأخطاء مع تلك الأخطاء التي فعلناها عندما قمنا بعملية التنبؤ وفقاً لخط انحدار أقل المربعات LSR. فهل هذا الخطأ فعلياً يحسن من توقعاتنا؟ انظر شكل (8).



شكل (9-8) خط الانحدار عندما يكون لدينا معرفة بالمتغير المستقل (X)

من هذا الشكل يمكننا أن نلاحظ أنه إذا كان هناك ارتباط وثيق بين هذين المتغيرين، وأن خط انحدار أقل المربعات سيقلل من معدل الخطأ. وبالتالي فإن الهوة بين نقاط هذه البيانات والخط ستكون قليلة جداً باستخدامنا لخط انحدار أقل المربعات أكثر منه باستخدام الخط الأفقي استناداً على فرضية لا يوجد ارتباط.

إن استخدام معامل التحديد كأحد مظاهر خط الانحدار هو الأساس في توضيح ذلك. فقيمة (r^2) تساوي 0.65 تشير إلى أن خط الانحدار لأقل المربعات Least Squares regressim Line تفسر 65% من التباين في المتغير التابع نسبة إلى التباين الذي تم تفسيره من خلال الخط الأفقي وهذه العملية تبين لنا التقليل الجوهري في معدل الخطأ(e).

يتعين علينا هنا أن نميز بين قيمة r و r^2 باعتبارهما قيمتان مرتبطتان بعضهما ببعض. ويعتبر معامل الارتباط مقياس معياري للعلاقة بين متغيرين؛ فمعامل الارتباط يُشير إلى المدى الذي يحدث فيه التغير في أحد المتغيرات يرتبط بالتغير في متغير آخر. عليه فإن قيمة r (تشبه b) التي هي في الأساس أداة للتنبؤ. أما على الجانب الآخر، فإن معامل التحديد يُعتبر مقياساً لتقليل النسبة في الخطأ (PRE)، في كمية التباين المفسر من خلال خط الانحدار، عليه فإن معامل التحديد يمد الباحث بإحساس كم لديه من الثقة يمكن الاعتماد عليها في صحة التنبؤ الذي يقوم به [(5)].

التباين الكلي: المفسر وغير المفسر:

يمكننا بشكل محدد تعريف التقليل في الخطأ الذي ينتج عندما تأخذ المعلومات حول المتغير المستقل في الاعتبار. هناك مجموعان مختلفان يمكن إيجادهما وبعد ذلك مقارنتهما بالتباين الكلي لِـ y لبناء إحصاء سيشير إلى التحسن في التنبؤ.

1- مجموع التباين المفسر Explained Variance ويوضح لنا مجموع التباين المفسر، يحسن قدرتنا على التنبؤ بـِ y عندما تؤخذ المعلومات المتعلقة بالمتغير المستقل (X) في الاعتبار. ويمكننا الحصول على هذا المجموع من خلال طرح $\bar{y}$ (تكهن درجة y بدون معلومات حول X) من الدرجات المتنبأ بها بواسطة معادلة الانحدار (درجة $\bar{y}$ أو y تم التنبؤ بها عندما كانت المعلومات متوفرة حول X). لكل حالة تربع وتجمع هذه الفروق. ويمكن تلخيص هذه العملية كالتالي: $\sum(\hat{y}-\bar{y})$. وأن النتيجة المتحصل عليها يمكن بعد ذلك مقارنتها بالتباين الكلي في y لنتحقق إلى أي مدى تكون معرفتنا بـِ X قد حسنت من قدرتنا على التنبؤ بـِ y.

والمعادلة المتعلقة بذلك يمكن بيانها رياضياً كالتالي:

التباين المفسر / التباين الكلي = $r^2 = \frac{\Sigma(\hat{y} - \overline{y})^2}{\Sigma(y - \overline{y})^2}$

إن معامل التحديد أو التقدير أو r^2 هي نسبة التباين الكلي في y الذي تم تفسيره من خلال X. وتشير r^2 بالتحديد إلى المدى الذي يساعد متغير X على التنبؤ لفهم أو تفسير متغير y.

2- مجموع التباين غير المفسر: لمعرفة التباين غير المفسر فإننا نقوم بطرح درجات $\hat{y}$ المتنبأ بها من درجات y الفعلية لكل حالة من الحالات؛ وبعد ذلك تربع وتجمع هذه الفروق. ويمكن تلخيص هذه العملية في الآتي:

$$\sum(y - \hat{y})^2$$

إن المجموع المتحصل عليه سيقيس كمية الخطأ في التنبؤ في y التي تبقى حتى بعد أن أخذنا متغير X في الاعتبار. إن نسبة التباين الكلي في y غير المفسرة لِـ X يمكن أن نتحصل عليها من خلال طرح قيمة r^2 من 1.00.

وتجدر الملاحظة هنا أنه كلما كانت العلاقة قوية بين X و y كانت قيمة التباين المفسر أكبر والعكس بالعكس. أنه في حالة العلاقة التامة $r = \pm 1.00$ فإن التباين غير المفسر سيكون صفر وأن r^2 ستكون 1.00 وهذا يشير إلى أن X تفسر كل التباين في y، وأننا بهذا بإمكاننا التنبؤ بقيمة y من X بدون خطأ (e). وعلى الجانب الآخر، فإذا كانت X و y لا يرتبطان بعضهما ببعض (r = 00.0)، فإن التباين المفسر سيكون (0) وأن r^2 ستكون كذلك 0.00 في مثل هذه الحالة، فإننا سنتوصل إلى نتيجة مفادها أن X لا يفسر أي شيء من التباين في y وأننا سنكون غير قادرين على التنبؤ بـِ y. ففي المثال الذي تعاملنا معه في هذا الفصل تحصلنا على r = 81 وبعد تربيع هذه الفئة تحصلنا على معامل تحديد تصل إلى 0.65. (r^2 = 0.65) والتي تشير إلى أن 65 % من التباين الكلي في حالات الاستياء الشعبي في هذه المدن تم تفسيره. ومن هنا نجد أن 35 % من التباين

الكلي في y لم يتم تفسيره من قِبل X، وقد يرجع السبب في ذلك إلى تأثير بعض المتغيرات المؤتلفة، أو إلى خطأ القياس، أو الصدفة العشوائية [6].

القياس والمتغيرات الديميوية Dummy Variobles:

يُعتبر معامل الارتباط والانحدار من أقوى التقنيات الإحصائية، لذلك فإنه في أغلب الأحيان تستخدم هذه التقنيات لتحليل العلاقة بين المتغيرات التي لم يتم قياسها على المستوى ذي المسافات / النسبي. إن هذه الممارسة بشكل عام، لا تمثل أية مشكلة إذا تعلق الأمر بالمتغيرات الترتيبية المتصلة التي لديها مدى واسع من الدرجات المحتملة، بالرغم من أن مثل هذه المتغيرات لا تمتلك نقاط الصفر الحقيقية، ومسافات متساوية من درجة إلى درجة. وقد بينا ذلك عند دراسة ارتباط سبيرمان "راهو" Spearmans` rho. كذلك يستخدم الباحثون في العلم الاجتماعي الارتباط والانحدار عند التعامل مع البيانات الترتيبية المضغوطة Collapsed. إن هذه المتغيرات تمتلك عدداً من الدرجات المحدودة (عادة بين 2 إلى 5)، مثل المسح الذي يبحث في الاتجاه نحو عقوبة الإعدام. إن مثل هذا الانتهاك لمستوى القياس لا يشكل بوضوح مشكلة طالما أن النتائج تعالج بدرجة مناسبة من الحذر.

إن هذه المرونة لدى الباحثين في العلم الاجتماعي ينبغي ألا تتسع لتشمل المتغيرات المقاسة على المستوى الاسمي. إن حساب معامل الارتباط أو الانحدار لمتغيرات مثل الحالة الاجتماعية، أو الطائفة الدينية تعتبر غير ذات جدوى؛ وذلك لأن درجات المتغيرات الاسمية ليست أرقام ولا تمتلك أي خاصية رياضية. إنه بإمكاننا أن نخصص درجة للديانة، الإسلام بدرجة (2) وغير الإسلام بدرجة (1). ولكن الدرجة المتحصلة للديانة الإسلام ليس ضعف الدرجة التي تم تحصيلها للديانة غير الإسلام. وبالتالي فإن الدرجات المتعلقة بالمتغيرات المقاسة على المستوى الاسمي هي مجرد عبارات وليست أرقام.

ولما كانت هذه المتغيرات متغيرات غير رياضية، فإنه من غير المنطق أن نقوم بحساب ميل الخط (b) أو نناقش ما إذا كانت هناك علاقات موجبة أو سالبة.

ومن هنا نجد أن الكثير من المتغيرات المهمة في حياتنا اليومية مثل: النوع، الحالة الاجتماعية، العرق، هي متغيرات اسمية لا يمكن تضمينها في معادلة الانحدار أو التحليل العلائقي، تلك التقنيات الأرقى والأقوى المؤثرة في مجال بحوث العلم الاجتماعي.

تجدر الإشارة إلى أنه من حسن الطالع أن الباحثين قد طوّروا طريقةً لمعالجة هذه المعضلة لكي نضيف المتغيرات المقاسة على المستوى الاسمي وذلك من خلال خلق ما يسمى بالمتغيرات الديميوية Dummy Variables. إن المتغيرات الديميوية يمكن أن تكون على أي مستوى من مستويات القياس بما فيها المتغيرات الاسمية، وإن لديها بالضبط فئتين، الأولى ترمز كـ: (0) والأخرى كـ: (1) وبالتالي ينبغي التعامل بهذه الكيفية مع المتغير المقاس على المستوى الاسمي لكي يكون ذا معني عندما يقدم كمتغير ثنائي يمكن تضمينه لمعادلة الانحدار. على سبيل المثال، يمكننا عند التعامل مع متغير النوع أن يخصص الرمز (0) للذكور، ويخصص الرمز (1) للإناث، العرق: أوروبي (0)، أسيوي (1) (ربما يحتاج الباحث إلى متغيرات ديميوية لإضافة أعراق أخرى أو انتماءات أخرى)؛ الانتماء الديني الإسلام قد يعطى له رمز (2) والمسيحية رمز (0) (مرة ثانية قد يحتاج الباحث إلى متغيرات ديميوية لأخذها في الحسبان لدى ديانات أخرى).

ولتوضيح ذلك، نفترض أننا نقوم بدراسة حول العلاقة بين العرق والتعليم (يقاس التعليم بعدد السنوات التي تم إنجازها). فإذا تم تخصيص رمز (0) للشخص ذو العرق الأبيض، والرمز (1) للشخص ذو العرق الأسود، فإنه بإمكاننا حساب معادلة ميل الخط y . وتكتب هذه المعادلة مستخدمين العرق كمتغير مستقل وحينها نبحث في العلاقة بين هذين المتغيرين. مرة ثانية، نفترض أننا قمنا بقياس عينة مستوى التعليم لدى العرق الأسود وخصصنا رمز (1). للعرق الأسود، والرمز (0) للعرق الأبيض، فإننا سنتحصل على معادلة الانحدار التالية:

$$y = a + bX$$

$$y = (12.0) + (-0.5)X$$

إذا كان التعليم متغير تابع y، والعرق متغير مستقل (X)، فإن خط الانحدار يعبر

العمودي Vertical لشكل الانتشار للنقطة، حيث y=12.0 والقيمة b= −0.5 للميل وهي تشير إلى علاقة سالبة بمعنى أن المبحوث ذو الخلفية السوداء في هذه العينة يسجل متوسط أقل من السنوات الدراسية مقارنة بنظيره الأبيض. وأن النتيجة ستكون واحدة إذا ما تغيرت طريقة الترميز.

فإذا كان الرمز (1) قد أعطي للرجل ذو الأصول العرقية البيضاء، وصفر (0) للرجل ذو الأصول السوداء، فإن قيمة (b) ستبقى بالضبط كما هي، وذلك لأن عملية الترميز المتعلقة بالمتغيرات الديمِيوية هي عملية اعتباطية، كما هو الحال في المتغيرات الترتيبية. فالباحث يتطلب منه أن يكون واضحاً حول ما تشير إليه القيم المتعلقة بالمتغيرات الديميوية.

كذلك يمكننا أيضاً استخدام معامل بيرسون (r) لتقييم قوة واتجاه العلاقة بالمتغيرات الديميوية. فإذا تحصلنا على سبيل المثال، على قيمة r = −0.23 بين العرق والتعليم، فإننا بالتالي نستطيع القول بأنه توجد علاقة سالبة ضعيفة إلى متوسطة بين هذين المتغيرين لهذه العينة. وباتساق الإشارة المتعلقة بميل الخط (b)، يمكننا أيضاً القول بأن التعليم يقل كلما زادت العرقية أي التحرك من العرق الأبيض إلى العرق السود. كذلك في هذا المقام، يمكننا استخدام معامل التحديد، وذلك بالقول، بأن العرق يفسر 5 % ($r^2 = 0.23^2 = 0.05$) من التباين في التعليم [7].

إجراءات الرسم البياني، والارتباط والانحدار باستخدام SPSS:

أولاً: توليد الرسم الانتشاري Scatter Plot مع خط الانحدار regression Line:

1- من القائمة الموجودة في الجزء العلوي من الشاشة أختر:

Graphs ⟸ Scatter (يعطيك صندوق الحوار Scatter plot).

2- انقر على Define (يعطيك مربع Simple scatter plot).

3- انقر على Number of civil disturbances.

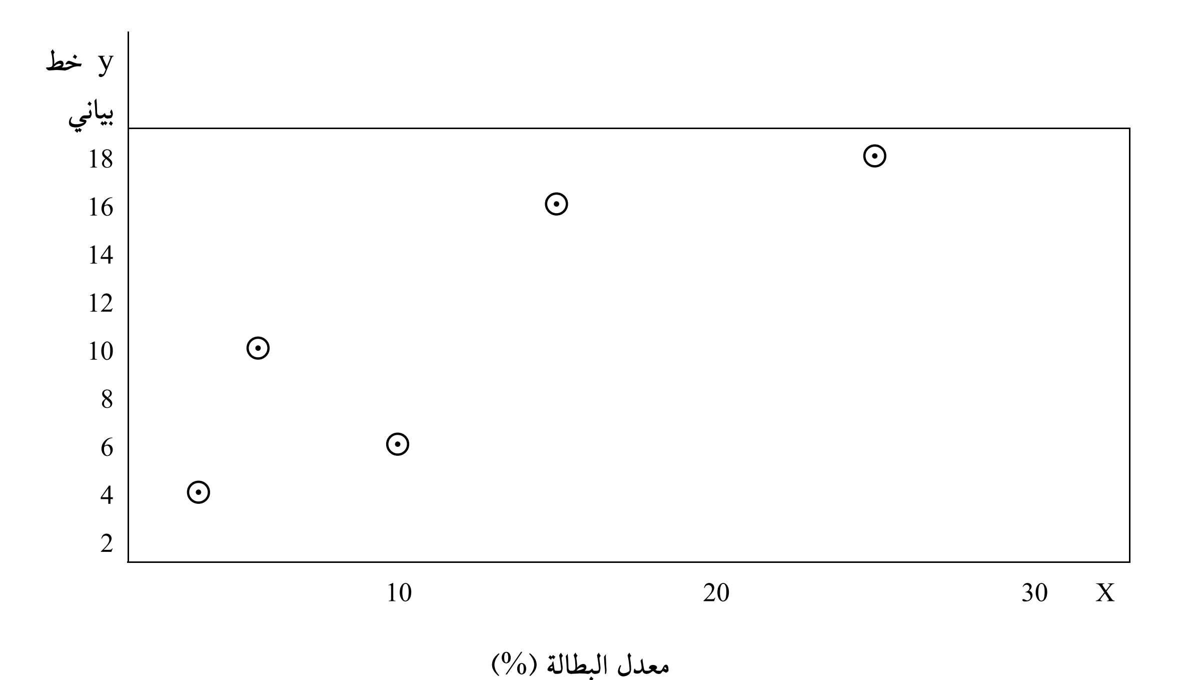
4- انقر على ◄ الذي يشير إلى y Axis في القائمة المحددة للمتغير (هذه العملية ستقود إلى لصق Number of civil disturbances كمتغير تابع يمكن وضعه على y Axis).

5- انقر على unemployment rate (معدل البطالة).

6- انقر على ◄ الذي يشير إلى X Axis في القائمة المحددة للمتغير (هذه العملية ستقود إلى لصق un employment rate كمتغير مستقل يمكن وضعه على X Axis في القائمة المحددة للمتغير المستقل).

7- انقر على OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء:



شكل (9 ـ 9) مخرجات SPSS للرسم الانتشاري

إجراءات إضافة خط الانحدار للرسم البياني باستخدام SPSS:

1- انقر نقرتان على الرسم البياني Scatter plot في نافذة Viewer. هذه العملية تقود إلى الرسم البياني في نافذة Chart Editor Window.

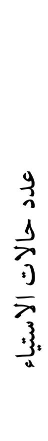
2- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

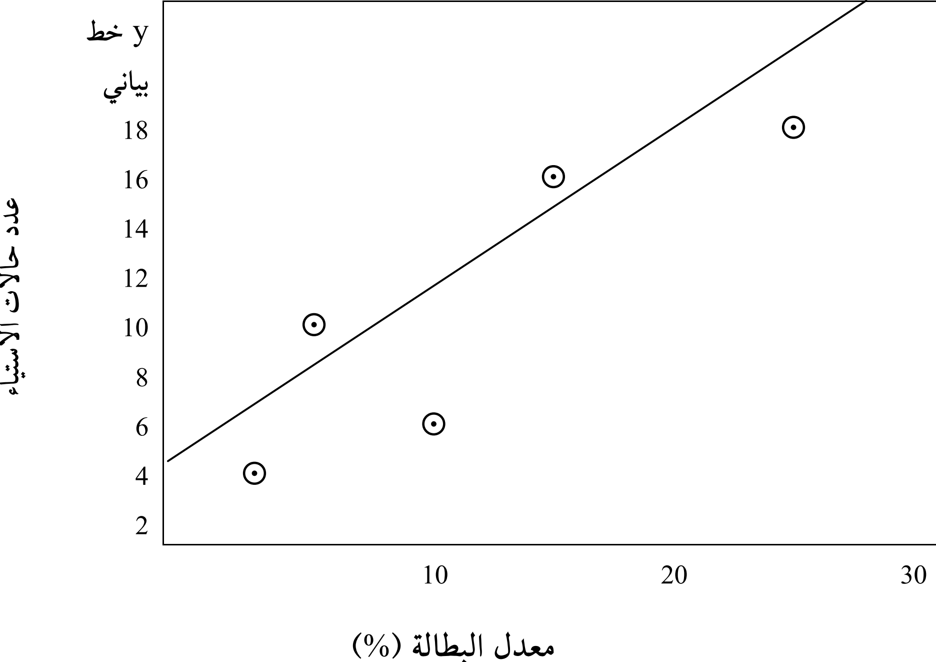
Chart ⟸ Options (هذه العملية تقود إلى Scatter plot options ومن خلال ذلك يمكنك رؤية المساحة المعنونة بـِ: Fit Line التي لديها خيار معنون بـِ Total بعلامة ✓ بجانبه).

3- انقر على علامة ✓ بجانب الـِ Total. هذه العملية تضع علامة ✓ في المربع لبيان أنه تم الاختيار.

4- انقر على OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء:





شكل (9ـ10) مخرجات SPSS للرسم الانتشار مع خط الانحدار

إجراءات الانحدار Regression مع تقدير المنحنى Curve Estimation باستخدام برنامج SPSS:

1- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Analyze ⇐ Regression ⇐ Linear

(تقود هذه العملية إلى مربع الحوار Linear Regression dialog box).

2- انقر على Number of civil disturbances.

3- انقر على ◄ التي تشير إلى Dependent في القائمة المحددة للمتغير (هذه العملية ستقود إلى عملية لصق Number of Civil disturbances كمتغير تابع).

4- انقر على Unemployment rate (متغير مستقل).

5- انقر على ◄التي تشير إلى Indepedent(s) في القائمة المحددة للمتغير (هذه العملية ستقود إلى عملية لصق Unemployment rate كمتغير مستقل).

6- انقر على OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

Regression

Variables Entered / Removed[b]

Model	Variable Entered	Variable Removed	Method
1	Unemployment Rate[a]		Enter

a. All requested Variables Entered

b. Dependent Variable: Number of Civil disturbances

Model Summery

Model	R	R Square	Adjusted R Squared	Std. Error of Estimate
1	.807[a]	.651	5.34	3.96

a. Predictors (Constant), unemployment Rate

ANOVA[b]

Model	Sum of squares	Df	MEAN square	F	Sig
Regression	87.701	1	37.701	5.586	.099[a]
Residual	47.099	3	15.700		
Total	134.800	4			

a. Predictors (Constant), unemployment Rate
b. Dependent Variable: Number of Civil disturbances

Correlation[a]

Model	Unstandardized coefficents		standardized coefficents	t	Sig
	B	Std Error	Beta		
(constant)1	4.423	3.013	.807	1.465	.239
Unemployment rate	.525	.222		2.364	.099

a. Dependent Variable: Number of Civil disturbances

المصدر: George Argyrous, op.cit, p. 218

شكل رقم (9 ـ 11) مخرجات SPSS للانحدار

تفسير مخرجات Spss للانحدار:

في مربع ملخص The Model Summery، نجد أن قيمة معامل ارتباط بيرسون (r) تساوي 807. ومعامل التحديد R square تساوي 651. وهاتان القيمتان هما نفس القيمتين اللتين تم حسابهما يدوياً. إن الجزء المهم من مربع معامل الارتباط Coefticients هو العمود المعنون بِـ B تحت معامل اللامعيارية Unstandardized coefficents ويعطينا هذا العمود المعلومات التالية:

- القيمة المتعلقة بِـ y-Intercept (والتي يطلق عليها a في التحليل أعلاه). إلا أن مخرجات SPSS يُطلق عليها كمية ثابتة Constant. وهي 4.423.

- ميل خط الانحدار The Slope of the regression Line الذي يشير إلى معامل الارتباط بمعدل البطالة 525.. مرة ثانية إن هذه النتائج هي نفسها التي تحصلنا عليها عند قيامنا بالعمليات الحسابية اليدوية.

اختبار t لمعامل الارتباط:

في الجزء الأول من هذا الفصل قد تم حساب خط الانحدار، ومعامل الارتباط الإحصائية المرتبطة بمجموعة من الحالات التي تم قياسها على المستوى ذي المسافات والنسبي. وقد تم التعامل مع هذه الإحصاءات الوصفية في إطار استقصاء العلاقة بين معدلات البطالة وحالات الاستياء الشعبي عبر مجموعة من المدن. والنتيجة التي توصلنا إليها كانت كالتالي:

(معدل البطالة) 0.53 + 4.4 = الاستياء الشعبي.

ومن خلال هذه الإحصاءات يمكننا الوصول إلى نتيجة مفادها إلى أن هناك تطابقاً موجباً وقويًّاً بين حالات الاستياء ومعدلات البطالة غير أنه يمكن القول، بأن هذه النتيجة المتحصل عليها من خلال عينة الدراسة قد لا تعكس ما يحدث في كل هذه المدن الخمس.

وكما هو الحال مع أي من الإحصاءات الوصفية الأخرى يمكن حسابها من عينة ما، فإننا بحاجة إلى أن نقرر ما إذا كان معامل الارتباط الذي يصف بيانات هذه العينة يعكس المجتمع الذي سحبت منه هذه العينة. فقد لا يكون هناك ارتباط بين هذين المتغيرين في المجتمع المكون لهذه المدن الخمس (ru)، وإنما يوجد فقط خطأ المعاينة الذي أدى بنا إلى اختيار هذه المدن الخمس التي قد لا تشبه باقي المدن الأخرى. وعليه، فإننا نحتاج إلى إجراء اختبار استدلالي لقيمة معامل الارتباط الذي تحصلنا عليه (r=0.81)[8].

في اختبار t لمعامل ارتباط بيرسون (r) يصاغ الفرض الصفري لهذا الاختبار: إنه لا يوجد ارتباط في المجتمع. في حين يصاغ الفرض البديل، إنه يوجد بعض الارتباط في المجتمع:

$$Ho : ru = 0$$

$$Hi : ru \neq 0$$

وبشكل واضح، فإن معامل ارتباط العينة قد وصل إلى: r = 0.81 وهي قيمة غير متطابقة مع الفرض الصفري. لكن السؤال المطروح هنا هو: هل باستطاعتنا رفض الفرض الصفري الذي مفاده: لا ارتباط في المجتمع على أساس نتيجة هذه العينة؟ ما هي الاحتمالية للحصول على عينة للمدن الخمس بارتباط يصل إلى r = 0.81 بين الاستياء الشعبي ومعدل البطالة من مجتمع يكون الارتباط فيه يساوي صفراً (0). وللحصول على هذه الاحتمالية، ينبغي علينا إجراء اختبار t مستخدمين المعادلة التالية:

$$t = \frac{r - ru}{sr}$$ العينة

$$S_r = \sqrt{\frac{1-r^2}{n-2}}$$

وبالتعويض للقيم المرتبطة بـِ: r و r^2 لهذه المعادلة، فإننا نتحصل على:

$$S_r = \sqrt{\frac{1-r^2}{n-2}} = \sqrt{\frac{1-(0.81)^2}{5-2}}$$

$$= 0.34$$

$$t = \frac{0.81-0}{0.34}$$

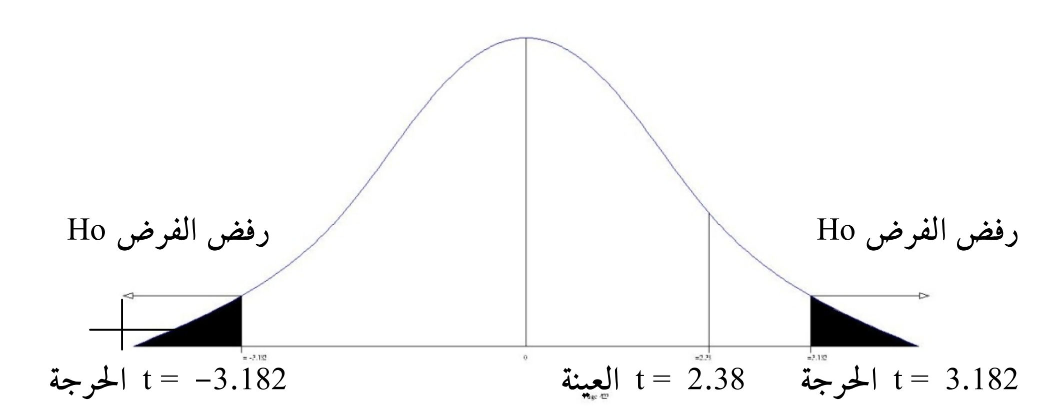
$$= 2.38$$

ولتحديد القيمة الحرجة لـِ t يمكننا الرجوع أو الاسترشاد بالجدول الذي يحتوي على القيم الحرجة لـِ t حيث إِنَّ درجة الحرية في هذا المثال تساوي N−2.

	مستوى الدلالة لاختبار أحادي الجانب				
	0.10	0.05	0.02	0.01	0.005
	مستوى الدلالة لاختبار ثنائي الجانب				
	0.20	0.10	0.05	0.02	0.01
df					
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
.	.....	.....	.....	.....	.....
.	.....	.....	.....	.....	.....
.	.....	.....	.....	.....	.....
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.011
14	1.345	1.761	2.145	2.624	2.917
15	1.341	1.753	2.131	2.602	2.947
16	.....	.....	.....	.....	.....
∞	1.282	1.645	1.960	2.326	2.576

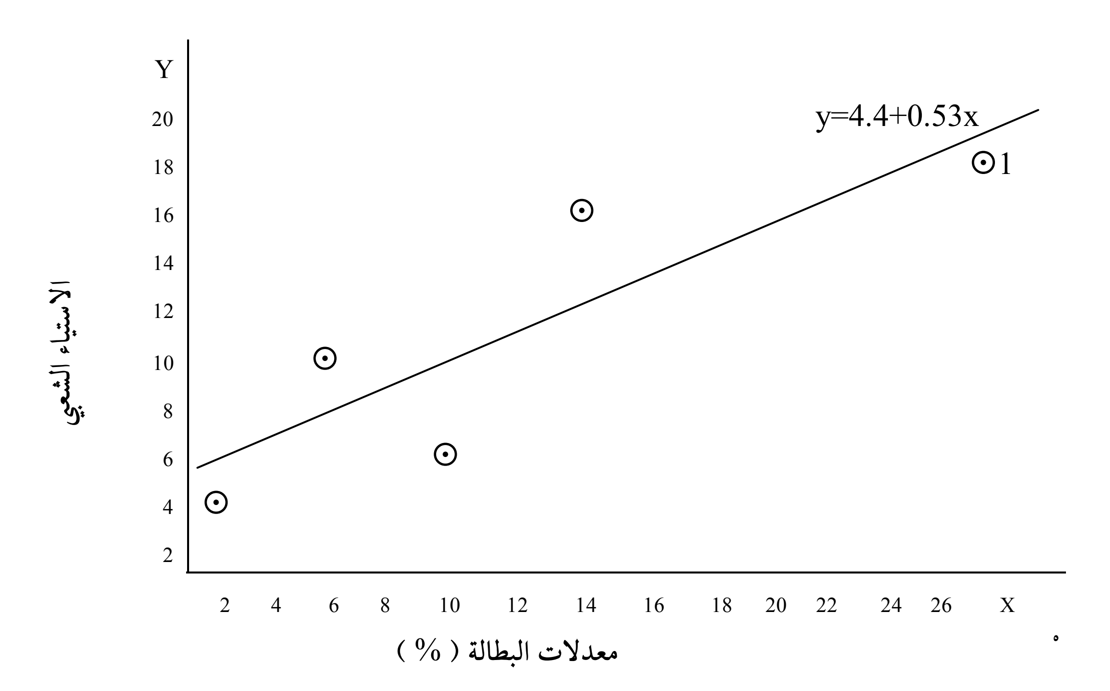
بمستوى ألفا 0.05 واختبار ثنائي الجانب، فإن درجة t الحرجة تكون كالتالي:
الحرجة t=3.182 ($\alpha = 0.05 , df = 3$)

وإذا وضعنا هذه المعلومات في شكل بياني (انظر الشكل رقم 11). فإننا بذلك نكون غير قادرين على رفض الفرض الصفري الذي مفاده لا يوجد ارتباط.



شكل (9-12) الدرجات الحرجة ودرجات العينة

إنه من الأهمية بمكان أن نتوقف لنفكر فيما حدث. ففي العينة التي بين أيدينا تحصلنا على ارتباط قوي بين المتغيرين X و y أي بين معدلات البطالة وحالات الاستياء الشعبي، ولقد أوضح لنا الاختبار الاستدلالي، أنه بالرغم من نتيجة هذه العينة التي ربما قد جاءت نتيجة لعامل الصدفة عند عملية المعاينة من المجتمع حيث إنّ هذين المتغيرين غير مرتبطين. ولنرى لِمَ لَمْ نستطع الإقرار بأن نتيجة العينة تعكس لنا هذه العلاقة في المجتمع؟ فإنه يصبح من الضرورة بمكان أن ننظر إلى شكل الانتشار لهذه البيانات مرة أخرى.



شكل (9ـ13) خط الانحدار لأقل المربعات (LSR)

يمكننا ملاحظة أن خط الانحدار قد تأثر بشكل كبير بالدرجة المتعلقة بمدينة رقم (1) التي يصل معدل البطالة فيها إلى 25 % يقابلها 17 حالة استياء شعبي. ولما كنا نتعامل مع عينة صغيرة (N=5)، فإن الحالة المتطرفة لاشك أنها تؤثر في النتيجة الكلية للعينة. فإذا كانت هذه الدرجة الواحدة مختلفة، كذلك خط الانحدار سيكون هو الآخر مختلفاً. وبما أن هذا الأمر محتمل، فمن غير الممكن أن يكون الارتباط القوي ذا دلالة عند التعامل مع العينات الصغيرة جداً [9].

اختبار الدلالة لارتباط بيرسون (r) باستخدام SPSS:

إجراء اختبار t لمعامل الارتباط الثنائي باستخدام SPSS:

1- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Analyze ⇐ Correlate ⇐ Bivariate

2- انقر على Number of civil disturbances.

3- انقر على ◄ مشيراً إلى المتغيرات في القائمة المحددة variable(s) هذه العملية تقوم بلصق Number of Civil disturbances في القائمة المحددة للمتغيرات (variable(s)).

4- انقر على Unemployment rate.

5- انقر على ◄ مشيراً إلى القائمة المحددة للمتغيرات variable(s) (هذه العملية تقود إلى لصق Unemployment rate في القائمة المحددة للمتغيرات(variable(s)).

6- انقر على OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

Correlations

Correlations

		Unemploment rate	Number of civil disturbances
Unemploment rate	Pearson Correlation sig.(2-tailed) N	1.000 5	.807 5
Number of civil disturbances	Pearson Correlation sig.(2-tailed) N	.807 5	1.000 5

المصدر: George Argyrous, op.cit, p. 425

شكل (9-14) مخرجات الارتباط الثنائي

تفسير اختبار الدلالة لمعامل ارتباط بيرسون (r) من خلال مخرجات SPSS :

إن اختبار الدلالة لارتباط بيرسون (r) يتولد كجزء من المخرجات عندما يجري تحليل الانحدار. إن الإجراء الذي تم إتباعه في الجزء الأول من هذا الفصل لتوليد إحصاءات الانحدار يولد لنا المعلومات الضرورية لإجراء اختبار الدلالة لهذه الإحصاءات.

تجدر الإشارة إلى أن هناك طرق أخرى بديلة من خلالها يمكننا توليد معامل الارتباط بين متغيرين وارتباط درجة t، ومستوى الدلالة.

ومن خلال أوامر الارتباطات بجزمة SPSS الثنائية Bivariate Correlation، يمكن للباحث توليد معامل الارتباط دون المعلومات الإضافية التي تأتي مع تحليل الانحدار الكامل[10].

من خلال النظر إلى مربع مخرجات الارتباط الثنائي يتبين لنا الارتباط بين المتغيرين. أولهما معدل البطالة مع نفسه. وثانيهما عدد حالات الاستياء الشعبي مع نفسه أيضاً. وكل واحد من هذين المتغيرين قد ولدّا معامل ارتباط 1.000. وهذه المعلومات تعتبر ضرورية طالما أن أي متغير بالضرورة يرتبط مع نفسه.

ففي الصف الأول من المربع يوضح لنا الارتباط بين معدل البطالة وحالات الاستياء الشعبي (807.)، وأن دلالة هذا الارتباط تصل إلى 099. وهذا يشير إلى أنه بالرغم من أن معامل الارتباط عالٍ، فهو ليس دالاً على مستوى (05.). وعليه يمكننا القول بأن هذه النتيجة جاءت كنتيجة لتباين المعاينة. أما الصف الثاني من المربع فيعطينا معامل الارتباط بين عدد حالات الاستياء الشعبي ومعدل البطالة وهو نفس المعامل المبين في الصف الأول من المربع.

اختبار t لمعامل ارتباط سبيرمان rho :

من خلال المثال السابق تعرفنا على الإجراءات المتعلقة بحساب معامل ارتباط بيرسون (r) ببحث ارتباط بيرسون (r) في استقصاء التطابق بين متغيرين تم قياسهما على المستوى ذي المسافات والنسبي. أما فيما يتعلق بارتباط سبيرمان، فإن هذا الإجراء يكون

ملائماً عندما يكون أحد المتغيرات على الأقل تم قياسه على المستوى الترتيبي. وإذا ما نظرنا بشكل دقيق إلى الإجراءات المتبعة في حساب هاذين النمطين من معامل الارتباط، فإننا نجد أنهما تقريباً متماثلان في الإجراءات، إلا أن الفرق بينهما هو أن معامل ارتباط (r) يمكن حسابه من البيانات الخام، في حين يعتمد ارتباط الرتب في حسابه على البيانات المرتبة.

تجدر الإشارة إلى أن اختبار الدلالة لكلا النمطين هو اختبار واحد، بمعنى، أن المعادلة لحساب درجة t للعينة هي واحدة بغض النظر، عما إذا كنا نختبر دلالة معامل ارتباط بيرسون (r) أو معامل الرتب $r_S(rho)$.

$$t = \frac{r_S - P}{S_r} \quad \text{العينة}$$

$$S_r = \sqrt{\frac{1 - r_S^2}{n-2}}$$

إذا أراد الباحث حساب قيمة $r_S(rho)$ فعليه أولاً حساب الفرق في الرتب لكل حالة من الحالات المدروسة، (D)، وبعد ذلك تربيع هذه الفروق. وبعد إنهاء هذه العمليات الحسابية ينبغي عليه تعويض تلك العمليات الحسابية في معادلة سبيرمان راهو (rho).

$$r_S = 1 - \frac{6\Sigma D^2}{n(n^2 - 1)}$$

فإذا توصلنا إلى ارتباط بين متغيرين يصل إلى $r_s = -0.8$ من بيانات تحتوي على $\sum D^2 = 1225.5$ لعدد 16 حالة.

فإننا بذلك يمكننا اتباع الإجراءات الخمسة المتعلقة باختبار الفروض:

الخطوة الأولى: صياغة الفرض الصفري والفرض البديل:

Ho: لا يوجد ارتباط بين X و y

$$Ho: P = 0$$

Hi: يوجد ارتباط بين X و y

$$Ho: P \neq 0$$

الخطوة الثانية: اختيار اختبار الدلالة:

حيث إننا نبحث في الارتباط بين متغيرين تم قياسهما على المستوى الترتيبي. وإن البيانات تم وصفها من خلال حساب ارتباط سبيرمان rho للرتب. فإن الاستدلال المناسب هو اختبار t لمعامل الارتباط.

الخطوة الثالثة: حساب درجة العينة:

لقد تم حساب معامل ارتباط سبيرمان راهو بين المتغيرين X و y لعدد 16 حالة، وجاءت قيمة الارتباط $r_S = -0.80$

$$r_S = \frac{6\sum D^2}{n(n_2 - 2)} = \frac{1225.5}{16(16^2 - 2)}$$

$$= -0.80$$

ولكي نرى ما إذا كانت هذه النتيجة قد جاءت للتباين العشوائي random variation عند إجراء عملية المعاينة من مجتمع حيث إنّ هذين المتغيرين لا يرتبطان بعضهما ببعض، فإننا بالتالي نحتاج أولاً إلى حساب الخطأ المعياري لتوزيع المعاينة لِ rho.

$$S_r = \sqrt{\frac{1 - r_S^2}{n - 2} = 1 - \frac{(-0.80)^2}{16 - 2}} \ .$$

$$= 0.16$$

وعليه، فإن درجة t للعينة ستكون كالتالي:

$$\text{العينة} \quad t = \frac{r_S - P}{S_r} = \frac{-0.80}{0.16}$$

$$= -5$$

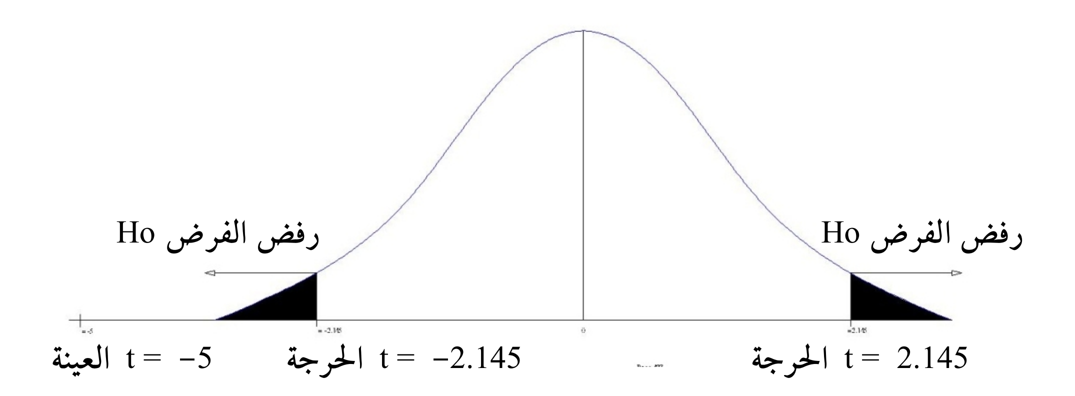
الخطوة الرابعة: حساب الدرجة (الدرجات) والمنطقة الحرجة:

بدلالة 0.05 = ∝ لاختبار ثنائي الجانب، يمكننا الإشارة إلى جدول توزيع t للقيم الحرجة لكي تقرر القيمة الحرجة لـِ t بدرجة حرجة 14 = 2 − 16 = df .

$$t = \pm 2.145 \ (\propto = 0.05 \, , df = 14)$$

الخطوة الخامسة: اتخاذ القرار:

إذا ما قمنا برسم درجات العينة والدرجات الحرجة في الشكل البياني التالي، فإنه يمكننا رفض الفرض الصفري. وعلى أساس هذه العلاقة القوية التي تـحصلنا عليها من خلال العينة المدروسة، فإنه ليس بإمكاننا القول بأنه لا يوجد ارتباط في المجتمع [11].



شكل (9-15) درجة العينة والدرجة الحرجة

أسئلة للمراجعة:

1- لماذا ينبغي على الباحث رسم شكل الانتشار للبيانات قبل إجراء تحليل الانحدار؟

2- ماذا تشير a و b في تحليل الانحدار؟

3- من خلال معادلات الانحدار بين اتجاه العلاقة؟

$y = 30 + 42 X$

$y = 30 - 0.38 X$

$y = -0.5 + 0.38 X$

$y = -0.5$

$y = -0.5 X$

4- اشرح الفرق بين ارتباط بيرسون (r) ومعامل خط الانحدار (b).

5- من البيانات التالية ارسم شكل الانتشار؟

a) X	5	6	9	10	10	13	15	18	22	27
y	35	26	30	22	28	28	20	21	15	18

b) من خلال هذا الرسم الانتشاري، بين الإشارة التي تسبق المعامل (على سبيل المثال، علاقة موجبة أو علاقة سالبة).

6- ما هي توقعات أقل المربعات Least Sqnare لـِ y عندما تكون X = 12 = y = 27.165 - 0.15X.

7- إذا كانت معادلة خط الانحدار $y = 40 + 0.7X$ لخط انحدار تم رسمه لبيانات متعلقة بعدد سنوات الحياة المتوقعة Life expectancy ومصروفات الدولة على الرعاية الصحية لكل فرد من السكان لمجموعة من الدول النامية. وقد اعتمد عدد سنوات الحياة المتوقعة كمتغير تابع. ومصروفات الدولة على الرعاية الصحية كمتغير مستقل. وأن معادلة خط الانحدار كالتالي: $y = 40 + 0.7X$.

- ما هو عدد سنوات الحياة المتوقعة إذا قامت الدولة بصرف مبلغ قدره 30.000 دينار على الرعاية الصحية لكل فرد؟
- ما هو عدد سنوات الحياة المتوقع إذا لم تقم الدولة بصرف أية مبالغ مالية على الرعاية الصحية عندما تكون $y = 40$، $X = 0$.
- هل يمكنك القول بأنه توجد علاقة قوية بين المتغيرين؟

8- البيانات التالية تبين عدد الأطفال ومساهمة الزوج في أعمال المنزل (بيانات تصورية):

العائلة:	1	2	3	4	5	6	7	8	9	10	11	12
عدد الأطفال:	1	1	1	1	2	2	3	3	4	4	5	5
عدد الساعات الأسبوعية التي يقضيها الزوج في العمل المنزلي	1	2	3	5	3	1	5	0	6	3	7	4

- أوجد شكل الانتشار وماذا يعني ذلك؟
- أوجد قيمة a، b.
- ارسم معادلة خط الانحدار y على X.

9- مسح اجتماعي توصل إلى معامل ارتباط بين عدد السنوات الدراسية التي يقضيها المبحوث، واتجاهاته نحو تولي المرأة مهام إدارية عليا.

قيمة الارتباط (r) تساوي $r = 0.54$، وحجم العينة المدروسة في هذا المسح كانت 140 حالة، ودلالة هذا الارتباط كان اختبار t الذي وصلت قيمته إلى 7.54.

ما هي النتيجة التي يمكنك الوصول إليها حول طبيعة هذه العلاقة بين هذين المتغيرين؟

10- من البيانات التالية: ارسم شكل الانتشار ؟ أوجد معادلة خط الانحدار y على X. إذا كانت X = 16. ما هي القيمة التقديرية لـِ y.

$$b = 0.87, a = 3.775$$

$$\hat{y} = \text{---} + \text{---}\ X$$

X	18	14	10	15	7	12	13	8	9	17	15	12
y	20	11	14	16	10	10	17	11	12	20	18	12

الهوامش والمصادر:

أولا: الهوامش:

1- George Argyrous, Statistics for Social and Health Research, With a Guide to Spss Sage Publications ,London, 2001, P. 201.

2- Ibid, PP. 201 - 209.

3- سعد اللافي، الإحصاء الاستنتاجي، ج 1، منشورات أكاديمية الدراسات العليا، طرابلس، 2003، ص 379.

4- George Argyrous, op.cit., PP. 209 - 212.

5- Ibid, P. 213.

6- Joseph F. Healey, The Essentials of Statistics: A TOOL for Social Research, 2[ed], Wadsworth Cengage Learning, UAS, 2010, PP. 343 - 345.

7- Ibid, PP. 349 - 350 and George DiEkhoff, Statictics for the social and Behavioral Sciences: Univariate, Bivariate, Multivariate, the Megraw Hill, Companies, Inc. USA, 1992, P.279.

8- George Argyrous, op.cit., P. 421.

9- Ibid, P. 424.

10- Ibid, P. 424.

11- Ibid, PP. 425 - 428.

ثانيا: المصادر:

1- Frederick J. Gravetter and Larry B. Wallnau, Statistics for The Behavioral Sciences, 8th ed, USA, 2010.

2- George Argyrous, Statistics for Social and Health Research, With a Guide to Spss, Sage Publications ,London, 2001.

3- George DiEkhoff, Statictics for the social and Behavioral Sciences: Univariate, Bivariate, Multivariate, the MCGraw- Hill, Companies, Inc. USA, 1992.

4- Joseph F. Healey, The Essentials of Statistics: A TOOL for Social Research, Wadsworth Cengage Learning, UAS, 2010.

5- سعد اللافي، الإحصاء الاستنتاجي، ج 1، منشورات أكاديمية الدراسات العليا، طرابلس، 2003.

الجزء الثالث

الإحصاءات الاستدلالية البارامترية

[حالة العينة الواحدة]

- الفصل العاشر: توزيعات المعاينة ـ التقدير، وفترات الثقة
- الفصل الحادي عشر: اختبار الفروض: اختبار Z لمتوسط عينة واحدة
- الفصل الثاني عشر: اختبار t لمتوسط حسابي لعينة واحدة

الفصل العاشر

توزيعات المعاينة - التقدير، وفترات الثقة

أولاً: توزيعات المعاينة:

تلعب الإحصاءات الوصفية دوراً كبيراً في تلخيص التوزيعات عموماً تلك التوزيعات التي تساعد في الإجابة على السؤال المطروح. فإذا كانت مجموعة الحالات التي تم أخذها كمقياس تشمل كل الحالات المحتملة التي نرغب في قياسها "المجتمع الإحصائي" فإن عملية البحث سوف تنتهي بالعملية الحسابية لهذه المقاييس الوصفية فالاستقصاء الذي يشمل كل عضو من السكان يُطلق عليه الإحصاء السكاني وأن الإحصاءات الوصفية لهذا المجتمع يُطلق عليها معلمات. فالمعلمة: هي ملخص وصفي لمتغير محدد في المجتمع فمتوسط الدخل لكل العائلات في مدينة بنغازي مثلاً هو معلمة. كذلك الأمر في التوزيع العمري لسكان مدينة بنغازي. وعندما يقوم الباحث بالتعميم من العينة فهو يستخدم ما يلاحظه من خلال العينة للتكهن بمعلمات المجتمع. وعند استخدام المعادلة الرياضية فإن المعلمات تشير إلى u كمتوسط للمجتمع و σ كانحراف معياري.

وتجدر الإشارة إلى أنه في بعض الأحيان تكون لدينا معلومات حول المجتمع ككل كالإحصاءات التي تجريها دولة ما ومن خلال هذه الإحصاءات يمكننا معرفة التوزيع

العمري لكل السكان في فترة تاريخية محددة. إلا أننا في وقت آخر قد لا يكون في إمكاننا معرفة التوزيع العمري لمجتمع ما، بالرغم من وجود هذا التوزيع. وعليه، فإننا في تلك الحالة نقوم بإجراء بحوث في الغالب نتعامل مع جزء من المجتمع أي عينة. وتسمى المقاييس الوصفية التي تستخدم لتلخيص المعلومات المتعلقة بالعينة بإحصاء العينة: وهو الملخص الوصفي لمتغير في عينة تستخدم للتكهن بمعلمة المجتمع. ويشار إلى إحصاء العينة بالتعبير الرياضي ب: $\overline{X}$ الذي يشير إلى متوسط العينة و S للانحراف المعياري لهذه العينة.

توجد لدى الباحث عدة أسباب تدعو إلى سحب عينة من المجتمع بدلاً من إجراء المسح الشامل لكل مفردات المجتمع وهذه الأسباب:

- قد يكون من الصعوبة بمكان على الباحث أن يحصر كل أفراد المجتمع إما لعدم توفر قائمة تحتوي على كل مفردات المجتمع أو أن بعض أفراد المجتمع لا يرغبون في المشاركة في مثل هذه الدراسة.

- قد تكون المعاينة في بعض الأحيان أكثر دقة فإذا كان لدينا أي سبب أن نعتقد بأن عملية المسح عن طريق العينة تولد أخطاء عندئذٍ فإن إجراء الإحصاء الشامل يمكن أن يوسع من هذه الأخطاء. فعلى سبيل المثال، فإن الإحصاء الشامل يتطلب فريقاً بحثياً كبيراً لإجراء المسح يمكن أن يقود إلى جامعي بيانات ليست لديهم الخبرة في جمع البيانات، في حين أن الفريق الصغير يمكن أن يكون مدرباً تدريباً جيداً ولديه خبرة كبيرة في جمع البيانات، ومهما كانت هذه الأسباب، فإن المشكلة الأساسية التي تظهر هي: هل الإحصاءات الوصفية التي نتحصل عليها من خلال العينة مساوية للإحصاءات التي نتحصل عليها إذا ما قمنا بإحصاء دقيق وشامل. هل إحصاء العينة يكون ممثلاً للمجتمع الذي سحبت منه هذه العينة؟

لاشك أن الباحث تكون لديه القدرة على سحب عينة ممثلة من المجتمع وأن عملية التباين العشوائي قد تؤثر في العينة من حيث التمثيل. والسؤال المطروح في هذا السياق هو: ما هي الأسس التي على ضوئها يمكننا أن نقوم بتعميم صادق من العينة على المجتمع؟ فعلى سبيل المثال، يمكننا أن نسحب عينة من 120 فرداً من منطقة جغرافية محددة

ونطرح على كل واحد منهم سؤالاً مفاده: كم يبلغ من العمر بالسنوات؟ ففي هذه الحالة يكون متغير العمر هو الأساس وقد تم قياسه على المستوى ذي المسافات والنسبي وأنه بإمكاننا أن نصف المعلومات التي تحتويها هذه البيانات من خلال حساب مقياس النـزعة المركزية لكي نتحصل على توزيع الدرجات، وبحساب مقياس التشتت يكون لدينا الإحساس بالدرجات حول المتوسط. ومن خلال الرسم البياني سيعطينا عموماً الانطباع حول التوزيع، إنه يمكننا الشعور بأن هذه الطرق ليست الوحيدة لوصف التوزيع كما بيّنا. لكن في الغالب قد نفي بمتطلبات الكثير من الأسئلة البحثية. إن هذه المعلومات قد تكون ذات أهمية في ذاتها إلا أننا قد نقوم بجمع بيانات حول عينة ما لأن لدينا سؤالاً آخرَ نود طرحه: ما هو المتوسط العمري لكل الناس في هذه المنطقة فإذا كان المتوسط العمري لهذه العينة هو 36 سنة فالسؤال الذي يمكن طرحه هو: هل بإمكاننا التعميم من هذه العينة على المجتمع ككل. هنا تظهر عملية التباين العشوائي التي قد يتسبب في إعاقة قدرتنا على التعميم الصادق من خلال إحصاء العينة.

كيف لنا أن نكون واثقين بأن العينة التي قمنا بسحبها لم يتم سحبها عن طريق الصدفة لتحتوي على نسبة قليلة من كبار السن أو نسبة قليلة من الصغار فيما يتعلق بالمجتمع. قد نطرح هذه المشكلة من خلال استخدامنا لإحصاء الاستدلال، فإحصاءات الاستدلال هي تقنيات عددية تهدف للوصول إلى نتائج حول المجتمع استناداً على البيانات التي تم الحصول عليها من عينة عشوائية تم سحبها من ذلك المجتمع. ولإجراء الإحصاء الاستدلالي فإنه يمكننا توليد ثلاث مجموعات عددية منفصلة:

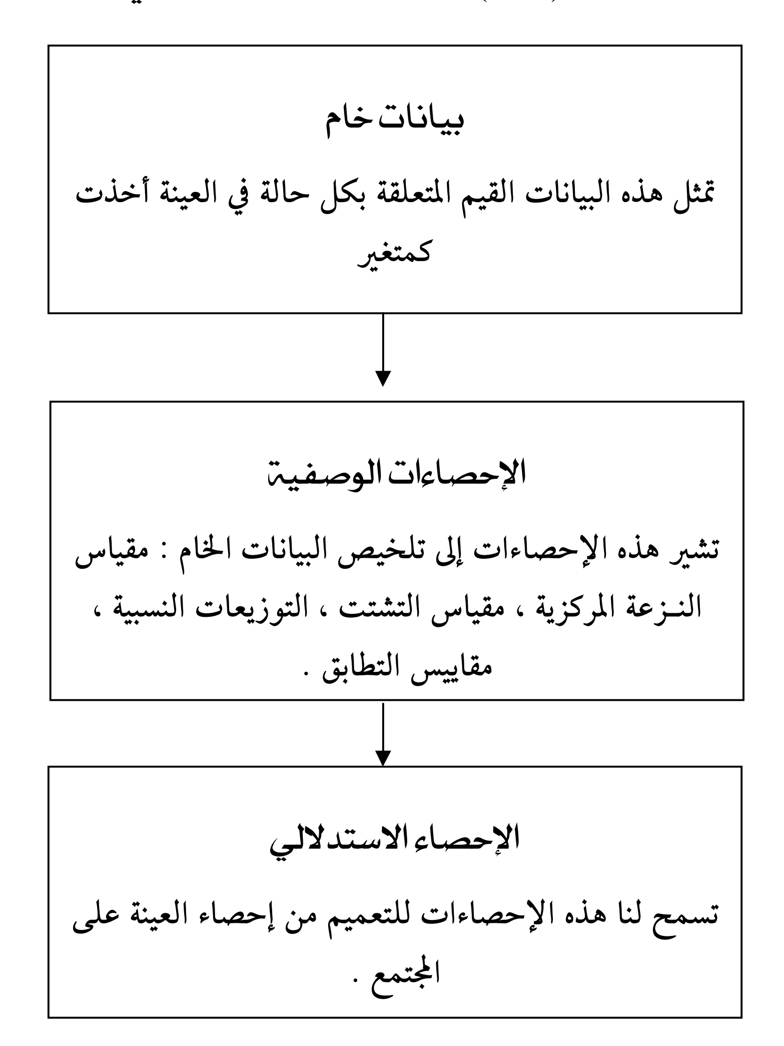
1- **بيانات خام:** وتمثل هذه البيانات المقاييس التي أخذت من كل حالة للمتغير. فالعمر على سبيل المثال تم قياسه بعدد السنوات. إن مثل هذه البيانات عادة ما تكون بيانات كثيرة اعتماداً على حجم العينة الفعلي.

2- **إحصاءات العينة:** وهذه تشير إلى الإحصاءات الوصفية التي تلخص البيانات الخام التي تم الحصول عليها من العينة (المتوسط، الانحراف المعياري أو التوزيع التكراري).

3- **الإحصاء الاستدلالي:** يساعدنا هذا النوع من الإحصاء للوصول إلى قرار حول

الخصائص المتعلقة بالمجتمع استناداً على إحصاءات العينة. بالرغم من التباين في تفاصيل الخطوات المتعلقة بإجراء الاستدلال من موقف إلى موقف آخر. إلا أننا نستخدم نفس الإجراء العام الذي يضمن توليد المجموعات العددية الثلاث. الشكل التالي يوضح هذا الإجراء.

شكل (10 ـ 1) عملية التحليل الاستدلالي



المصدر: George Argyrous , op.cit , p. 233

العينات العشوائية:

إن أهم مطلب يمكن مراعاته إذا أراد الباحث التعامل مع الإحصاءات الاستدلالية للتعميم من العينة على المجتمع هو أن يختار عينة عشوائية من المجتمع، فالاختيار العشوائي هو أسلوب المعاينة الذي يتيح لكل فرد في المجتمع أن تكون لديه نفس فرصة الاختيار ليكون عضواً في العينة. فالمسح عن طريق الهاتف على سبيل المثال قد لا يمثل البتة المعاينة العشوائية، باعتبار أن كثيراً من الناس لا يملكون جهاز الهاتف أثناء إجراء عملية المسح؛ وبالتالي فإن المسح لا يتضمن تلك الأسر التي لا تمتلك جهاز هاتف متيحة الفرصة لتلك الأسر التي تمتلك أكثر من جهاز هاتف التواجد ضمن أفراد العينة، ومن هنا قد يقع الباحث في عملية التحيز أي اختياره مجموعة من الناس بدلاً من أخرى.

المعاينة العشوائية الطبقية:

في كثير من الأحوال قد يلجأ الباحث - وفقاً لأسباب معينة - إلى التعامل مع العينات العشوائية الطبقية بدلاً من العينة العشوائية البسيطة، فالعينة العشوائية الطبقية تستخدم في مجتمع قابل للتقسيم الطبقي فكل طبقة هي جزء من المجتمع نشك في أنه مجتمع متماثل في إطار المتغير الذي نسعى للتعامل معه. ومن مميزات العينة الطبقية أنها تعطي درجة كبيرة من التمثيل من خلال تقليلها من احتمالية خطأ المعاينة.

تجدر الإشارة إلى أن أخطاء المعاينة يمكن التقليل منها من خلال عاملين أولهما: أن العينة الكبيرة تولد قليلاً خطأ معاينةٍ إذا ما قورنت بالعينات الصغيرة. كلما كان المجتمع متجانساً قَّلَّ خطأ المعاينة والعكس بالعكس. فإذا وافق 99 % من السكان حول قضية معينة فهذا يعني أن خطأ المعاينة في مثل هذه الحالة سيكون قليلاً. أما إذا تمت الموافقة من قِبل السكان لـِ 50 % مقابل 50 % حول القضية المطروحة ففي هذه الحالة نتوقع خطأً معاينةٍ أكبرَ. ومن هنا نجد أن العينة الطبقية تستند في الأساس على العامل الثاني من نظرية المعاينة، ولكي نتحصل على عينة ممثلة لطلاب الجامعة على سبيل المثال، ينبغي أولاً أن نتحصل على قائمة للطلاب من خلال الكليات، وبعد ذلك يمكننا سحب عينات ملائمة من الطلاب حسب السنوات أو حسب السنوات وحسب التخصص.

أما في العينة غير الطبقية فإن التمثيل من خلال الكليات فقط سيؤدي بالضرورة إلى خطأ المعاينة[1] وعادة ما نطلق على المعاينة العشوائية المعاينة الاحتمالية. وتجدر الإشارة هنا إلى أن هناك عدداً كبيراً من تقنيات المعاينة غير الاحتمالية (ليست عشوائية) وتعتبر معاينة الكرة الثلجية شكلاً من أشكال المعاينة العرضية. وهذا النوع من المعاينة يكون ملائماً عندما يكون هناك عددُ من الناس يصعب على الباحث تحديدهم أي مجتمعُ غيرُ موحدِ المعالمِ من حيث الحجم، والتركيبة. كذلك تزداد أهمية هذه التقنية عندما يكون لدينا مجتمعُ مغلقُ، وقوي الارتباط يصعب على الباحث الدخول فيه. مثل: فئة المشردين والمهاجرين، والمهاجرين غير الشرعيين إلى آخر ذلك. ومن خلال هذه التقنية يقوم الباحث بجمع البيانات من مجموعة محددة من المبحوثين التي تمكن الباحث من الوصول إليها وعندما يقوم الباحث بطرح مجموعة من الأسئلة عليهم يطلب منهم الإدلاء بأشخاص آخرين يعرفونهم. إذاً معاينة الكرة الثلجية تشير إلى عملية تراكمية حيث إن كل شخص يتم مقابلته يمكن أن يشير إلى شخص أو أشخاص آخرين يعرفهم. ولما كانت المعاينة التلجية معاينة غير احتمالية فإنها أسلوب يمكن استخدامه بشكل فعال في الدراسات الاستكشافية[2].

وتجدر الإشارة في هذا السياق إلى أنه لا يوجد لدينا أي مبرر متأصل يجعلنا نفكر في أن المعاينة الاحتمالية هي أفضل من المعاينة غير الاحتمالية باعتبار أن كل طريقة من هذه الطرق تكون ملائمة لتساؤلات بحثية مختلفة. وفي بعض الأحيان يكون من الأفضل اختيار أسلوب المعاينة غير الاحتمالية للسؤال البحثي المطروح. وتجدر الإشارة هنا، إلى أن المهتمين والأكاديميين لا ينظرون دائماً بهذه الطريقة. فالبحث يتطلب نظرة علمية عند طرحه في إطار الإحصاء الاستدلالي.

وغالباً ما يندرج في إطار العمل ليتلاءم مع التقاليد العلمية. فالإحصاء الاستدلالي في بعض الأحيان يمكن حسابه من عينة لم يتم اختيارها عشوائياً، وفي أحوال أخرى، قد يبني مشروع البحث بطريقة ما تكون الإحصاءات الاستدلالية ملائمة لهذه الخطة بالرغم من أن هناك تقنياتٍ أخرى أكثر نفاداً. إن هذه هي مشكلة مرتبطة بالممارسة البحثية التي تثير قضايا أوسع لا يمكننا التعامل معها هنا. إلا أن الشيء الذي يمكننا أن ننوه عنه في

هذا الموضع، أن اختيار تقنية البحث يجب ألا تؤخذ على أساس التقنيات التي تستخدم في تحليل البيانات، وإنما اختيار تقنيات البحث يجب أن تكون على أساس أفضل طرح للمشكلة مثار البحث.

توزيع المعاينة لإحصاء العينة:

يُطَبَّقُ الإحصاءُ الاستدلالي فقط على العينات العشوائية باعتبار أن الأداة الأساسية لإجراء عملية الاستدلال تتوقف أساساً على المعاينة العشوائية ويُطْلَق على هذه الأداة توزيعُ المعاينة لإحصاء العينة. وقبل الدخول في تعريف توزيع المعاينة دعنا نبين فكرة خلق بناء توزيعات المعاينة من خلال العملية التالية: ففي المثال المتعلق بالمجتمع المحلي الذي تم مسحه وتسجيل الأعمار بالسنوات وأن معلمات هذا المجتمع البالغ عدده 400 مفردة تكون:

$$u = 35$$

$$\sigma = 13$$

دعنا الآن نفترض أننا لم نُجْرِ مسح الحالات البالغ عددها 400 شخصاً في هذا المجتمع المحلي وعوضاً عن ذلك قمنا بإجراء التجربة التالية:

أولا: اختيار 120 حالة عشوائية وَطُرِحَ عليهم السؤالُ المتعلقُ بالعمر وتم حساب متوسط أعمارهم.

ثانيا: إعادة هذه المجموعة إلى المجتمع المحلي الأصلي مرة ثانية وسحبت مجموعة أخرى تتكون من 120 شخصاً. (ربما هذه المجموعة تحتوي على أعضاء من العينة الأولى) ونستمر في سحب هذه العملية مرة تلو الأخرى، وفي كل مرة نقوم بحساب المتوسط العمري لكل عينة تم سحبها عشوائياً. والنتيجة النهائية لعدد 20 عينة عشوائية مكررة يوضحها الجدول التالي:

جدول (10 ـ 1) توزيع عشرين متوسط لعينات عشوائية

رقم العينة	متوسط العينة
1	34.7
2	35.9
3	35.5
4	34.7
5	34.5
6	35.4
7	35.7
8	34.6
9	37.4
10	35.3
11	34.1
12	35.5
13	34.9
14	36.2
15	25.6
16	35.0
17	35.1
18	36.4
19	35.6
20	33.6

من خلال بيانات هذا الجدول يتبين لنا أن معظم المتوسطات تتجمع حول قيمة المجتمع وهي 35 سنة مع وجود درجات قليلة تبتعد بعض الشيء عن قيمة متوسط المجتمع، وهي الدرجة المتطرفة 37.4 سنة. وقد جاء هذا المتوسط لهذه العينة عن طريق الصدفة خلال عملية التباين العشوائي لتشمل نسبياً مجموعة من الأعمار الكبيرة في المجتمع المحلي. وبالرغم من ذلك يجدر بنا أن نورد الحقيقة التي مفادها أن أعمار الأشخاص في هذا المجتمع تتراوح ما بين 2 و 69 سنة وبالتالي فإن متوسط العينات تحتوي على مدًى محدودٍ من القيم قرابة نصف عدد العينات العشرين التي تم سحبها والتي تولد متوسطاً عمريّاً داخل حدود نصف سنة من المتوسط الحقيقي للمجتمع وهذا يعطينا معنى بقيمة وثبات العينات العشوائية.

دعنا نتعمق أكثر في هذا المثال النظري ونتخيل نظرياً أننا قد أخذنا عدداً لا متناهي من العينات العشوائية متساوية في الحجم من هذا المجتمع ثم نلاحظ التوزيع لكل المتوسطات المتعلقة بهذه العينات. إن النمط الذي قمنا بمشاهدته أعلاه فإننا قد لاحظنا فقط أن عشرين عينة عشوائية سوف تُعزز. أن معظم العينات تتركز حول معلمات المجتمع مع وجود بعض العينات نسبياً تبتعد عن التوزيعات بشكل أو آخر. إن مثل هذا التوزيع يطلق عليه توزيع المعاينة. وتُعْرَفُ توزيعات المعاينة بأنها التوزيعات الاحتمالية النظرية الناتجة عن عدد لا متناهي من نتائج إحصاء العينة باستخدام عينات عشوائية متساوية الحجم، بمعنى آخر أن توزيعات المعاينة هي توزيعات نظرية تعتمد في بنائها على الممارسة المنطقية. كما يُعْرَفُ توزيع المعاينة: بأنها التوزيعات التي تصف ظهور قيم التوزيعات لبعض الإحصاء تم حسابها من كل أحجام العينات التي تم سحبها من بعض المجتمعات. وإذا ما نظرنا إلى مثل هذه التعريفات بشكل دقيق يمكننا القول:

1- توزيع المعاينة هو توزيع نظري لكينونات entities. ومع أن توزيعات المعاينة نظرياً ممكنة لسحب كل العينات الممكنة من المجتمع إلا أنه قد يتعذر من الناحية العملية على الأقل إذا كان حجم المجتمع كبيراً.

2- توزيعات المعاينة لا تكشف عن ظهور الدرجات التكرارية على بعض من المتغير ولكنها تصف بدلاً من ذلك لماذا في أحوال كثيرة نجد أن بعض القيم المتنوعة لبعض

الإحصاء تم الحصول عليه من خلال حساب العينات التي تم سحبها من المجتمع. على سبيل المثال. إن توزيعات المعاينة للمتوسط يتشكل من خلال سحب كل الأحجام الممكنة للعينة (N) من المجتمع وحساب المتوسط لكل عينة. كما يمكننا أيضاً أن نتخيل أننا قمنا بسحب عدد من الأحجام الممكنة للعينات (N) من المجتمع وقمنا بحساب التباين لكل عينة وهذا ما يطلق عليه توزيعات المعاينة للتباين.

3- إن توزيعات المعاينة ليست دائماً تستند على سحب كل العينات الممكنة من المجتمع فبعض توزيعات المعاينة يمكن الحصول عليها من خلال سحب كل الأزواج الممكنة من العينات، عينات ثلاثية أو عينات رباعية أو أي مجموعة متعددة من العينات ويحسب الإحصاء لكل زوج من العينات أو العينات الثلاثية أو الرباعية أو أي مجموعة متعددة من العينات تشكل بذلك هذه الحسابات الإحصائية توزيعات المعاينة[3].

إن توزيعات المعاينة للمتوسطات لها خصائص ثلاثة مهمة:

- إن متوسط توزيع المعاينة يكون مساوياً لمتوسط المجتمع بمعنى آخر أن متوسط المتوسطات ($u\overline{x}$) سيكون نفسه مثل متوسط المجتمع ويمكن التعبير عنه جبرياً بـ:

$$u\overline{x} = u$$

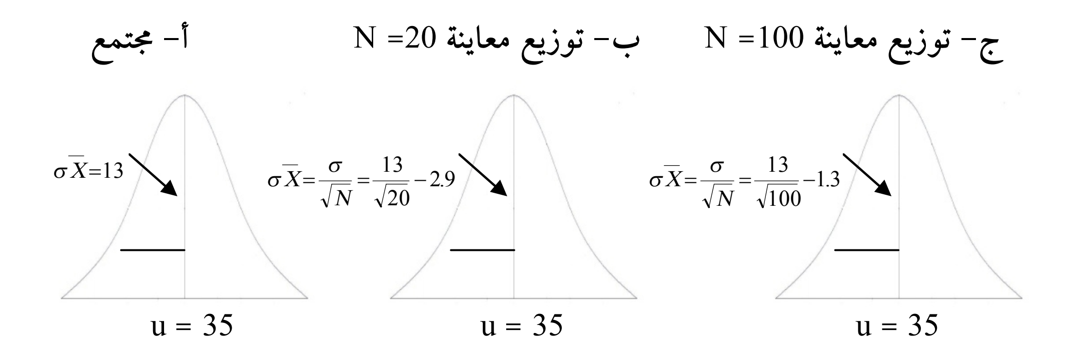
إن الخطأ المعياري سوف يرتبط بالانحراف المعياري للمجتمع. فالانحراف المعياري لتوزيع المعاينة يُعْرَفُ بالخطأ المعياري ($\sigma\overline{x}$) Standard error. وأن قيمته تتأثر بحجم العينة وكمية التباين في المجتمع. فإذا أخذنا على سبيل المثال العينة من خمسة أشخاص. وأن واحداً من هؤلاء الأشخاص في هذه العينة الصغيرة يبلغ من العمر 60 عاماً. فإن متوسط هذه العينة سيكون متأثراً بشكل كبير بهذه الدرجة. بمعنى آخر، أننا نتوقع أن تكون هذه العينة الصغيرة أقل ثباتاً إذا ما قورنت بالعينة الكبيرة، حيث من المتوقع أن تكون هناك احتمالية كبيرة للحصول على نتيجة واسعة من التشتت. أما إذا كان لدينا عينة حجمها 220 مفردة. فإن تأثير إحدى الدرجات الكبيرة سيضعف بتأثير مجموعة كبيرة من الحالات التي تكون قريبة من متوسط المجتمع. وعليه فإن إعادة تكرار عينات كبيرة ستتجمع قريباً من قيمة المجتمع؛ وبالتالي ستكون هذه العينات أكثر ثباتاً. وبشكل

مماثل إذا قمنا بسحب عينات من مجتمع تتوزع الأعمار فيه من سنتين إلى مئة واثنتين سنة، فإن مدى الدرجات التي سنتحصل عليها من هذه العينات ستكون أكبر مما لو قمنا بالمعاينة من مجتمع تتوزع فيه الأعمار بين 20 سنة و 30 سنة. إذاً كلما كان المجتمع متجانساً تجمعت العينات العشوائية بشكل وثيق حول ذلك المجتمع الذي سحبت منه هذه العينات. هذان العاملان يجذبان الانتباه من خلال المعادلة التالية للخطأ المعياري:

$$\sigma \overline{X} = \frac{\sigma}{\sqrt{}}$$

إن توزيعات المعاينة ستتوزع توزيعاً طبيعياً وأن نسبة العينات ستقع داخل مدى محدد من القيم التي ستحدد من خلال التوزيع الطبيعي المعياري.

إن الخصائص الثلاثة لتوزيعات المعاينة لمتوسطات العينة يوضحها الشكل التالي:



شكل (10 ـ 2) توزيعات المعاينة لعينات ذات أحجام مختلفة

من الشكل (10 – 2) يتضح أن شكل (أ) يبين توزيع الحجم الكلي للمجتمع المحلي البالغ عدده 400 نسمة. أما شكل (ب) فيشير إلى توزيع المعاينة لمتوسطات عينات يبلغ حجمها 20 مفردة. بمعنى آخر، هو توزيع المتوسطات التي سنتحصل عليها إذا ما كررنا سحب 20 مفردة من سكان هذا المجتمع. أما الشكل (ج) فهو توزيع المعاينة لمتوسطات عينة حجمها 100. من خلال هذه الأشكال يمكننا أن نرى أن كلا التوزيعين يتمركزان

عند متوسط المجتمع 35 سنة إضافة إلى أنهما موزعان توزيعاً طبيعياً، ومع ذلك فإن الخطأ المعياري لكل واحد من توزيعات المعاينة سيتباين. ومع تكرار عينات حجمها 20 فإن متوسطاتها ستتوزع بشكل كبير بخطأ معياري يصل إلى 2.9 سنة في حين أنه في العينات الكبيرة عندما كان حجم العينة 100 فإن نتائج هذه العينة تتجمع بشكل وثيق حول قيمة المجتمع وكلا التوزيعين موزعان توزيعاً طبيعياً، وبالتالي نجد أن 68 % لكل الحالات تقع خلال (1) انحراف معياري من المتوسط. في حين أن توزيع المعاينة عندما يكون حجم العينة 20 فإن هذا المدى سيكون بين 32.1 سنة و 37.9 سنة.

$$35 \pm 2.9 =$$

سنة 32.1

و

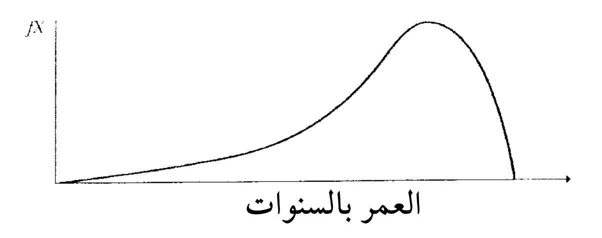
سنة 37.9

في حين نجد أن المدى في العينة الثانية لتوزيع المعاينة سيكون أكثر ضيقاً من توزيع المعاينة للعينة الأولى فالحد الأدنى 33.7 والحد الأعلى 36.3 سنة [4].

نظرية تقارب التوزيعات الاحتمالية:

قد تمت مناقشة خصائص توزيع المعاينة مستنتجة من مجتمع موزع توزيعاً طبيعياً. وعلى وجه الخصوص، فإن توزيع المعاينة هي الأخرى موزعة توزيعاً طبيعياً.

السؤال الذي يمكن طرحه هنا هو: ماذا لو أن أعمار الـ 400 نسمة لهذا المجتمع المحلي الصغير قد توزعت أعمارهم وفقاً للشكل التالي:



شكل (10-3) توزيع مائل

إن هذا التوزيع هو توزيع مائل نحو اليسار مشيراً إلى أن هناك نسبياً عدداً أكبر من كبار السن مقارنة بصغار السن بهذا المجتمع المحلي. وأن تكرار عينات عشوائية من هذا التوزيع المائل سيولد لنا كذلك. توزيعات معاينة مائلة أيضاً. ومع هذا فإن الأمر ليس كذلك. وطبقاً لإحدى القواعد الإحصائية الرئيسية: نظرية تقارب التوزيعات الاحتمالية Central Limit Theorem، أوضحت أن ثمة شروطاً معينة ستكون فيها توزيعات المعاينة مُوَزَّعُ توزيعاً طبيعياً، بالرغم من أن توزيع المجتمع الذي سحبت منه العينة موزعاً توزيعاً غير طبيعي. وتعرف نظرية تقارب التوزيعات الاحتمالية، بأنها عدد غير محدود من العينات العشوائية بأحجام متساوية تم اختيارها من مجتمع، وأن توزيع المعاينة لمتوسطات العينة سيقترب من التوزيع الطبيعي مثلما يقرب حجم العينة من اللامتناهي.

إن المجتمع يمكن أن تكون وحداته موزعة توزيعاً غير طبيعي، ومع ذلك فإن تكرار المعاينة (نظرياً) سيولد توزيعات معاينة طبيعية. في حقيقة الأمر، إن حجم العينة ليس بالضرورة أن يكون كبيراً كما تم افتراضه في الصيغة الأساسية لنظرية تقارب التوزيعات الاحتمالية: فعندما يزيد حجم العينة عن 100، فإن توزيع المعاينة لمتوسطات العينة سيكون على وجه التقريب طبيعياً [5].

خلاصة القول، بالرغم من النقاش المتعلق بتوزيعات المعاينة إلا أنه من الناحية العملية في مجال البحوث الاجتماعية والإنسانية عادة ما يلجأ الباحث إلى سحب عينة واحدة من مجتمع ومن ثم يقرر هل هذه العينة تعكس الخصائص العامة للمجتمع أم لا. إن السؤال الذي يمكن طرحه هنا هو: ما الفائدة من استخدام توزيعات المعاينة؟ إن توزيعات المعاينة تعتبر الحجر الأساس الذي يستند عليه الإحصاء الاستدلالي كما سنرى في الفصول اللاحقة.

ثانياً: التقدير وفترات الثقة:

في الجزء الأول من هذا الفصل حاولنا أن نبين الخصائص المتعلقة بتوزيع المعاينة لمتوسطات العينة. حيث يتميز توزيع المعاينة بثلاث خصائص مهمة:

1- إن متوسط توزيع المعاينة يكون مساوياً لمتوسط المجتمع، بالرغم من أن متوسط أي

وتجدر الإشارة إلى أنه يمكن افتراض أن أي زوج معطى من الآباء أو الأطفال يقع داخل نفس المجموعة الاجتماعية الاقتصادية. وبالتعامل مع زوج الآباء وزوج الأبناء، فإننا بالتالي ننظر إلى الفرق في كل زوج. وأن تأثير المتغيرات الأخرى كالوضع الاجتماعي والاقتصادي قد يزول Mitigated. في الواقع، نحن نقول إن كل المتغيرات الأخرى التي يمكن أن تحدد مشاهدة الإذاعة المرئية هي متساوية في كل زوج من الأزواج المعطاة، ولذلك، فإن العلاقة العائلية فقط تختلف بين هذه المتغيرات بحيث تسمح لنا بعزل تأثير هذه المتغيرات على المتغير التابع[10].

اختبار t لعينات تابعة لفرق المتوسط:

لتوضيح استخدام اختبار t لعينات تابعة (الأزواج) فإننا نورد المثال التالي:

مثال:

أُجْرِيَ مسحُ اجتماعيٌّ لعدد 10 عائلات، أب من كل أسرة، وطفل من كل أسرة، وطُلب إليهم أن يقوم كل منهم بتسجيل كمية المشاهدة للإذاعة المرئية خلال فترة معينة من الوقت في مدونة يومية. وقد ظهرت البيانات في الجدول التالي:

جدول (13ـ2): مشاهدة الإذاعة المرئية وفقا لزوج الأسر

الأسرة	الدقائق المشاهدة من قِبل الأب	الدقائق المشاهدة من قِبل الطفل
1	23	45
2	25	56
3	43	73
4	26	53
5	21	27
6	29	34
7	32	76
8	23	21
9	25	54
10	21	43
	$\overline{X}=\frac{\Sigma x}{N}=26.8$	$\overline{X}=\frac{\Sigma x}{N}=48.2$

المصدر: George Argyrous, Statistics for Social & Health Research, op.cit. , P. 435

ولإجراء اختبار t لعينات تابعة يمكننا قلب الإجراءات التي تم اتباعها في اختبار t لعينتين مستقلتين:

1- حساب المتوسط الحسابي لكل عينة.

2- وبعد ذلك حساب الفرق بين وسطي العينتين.

وعند قلب هاتين الخطوتين لإجراء اختبار t لعينتين تابعتين يصبح الإجراء كالتالي:

1- حساب الفرق لكل زوج من الحالات (D)

2- وبعد ذلك حساب متوسط الفروق ($\overline{X}_D$)

ولكي نضع هذا الأمر بشكل أكثر دقة، فإنه يمكننا القول بأن اختبار t للعينات المستقلة تعبر عن الفرق بين المتوسطات، بينما يعبر اختبار t للعينات غير المستقلة عن متوسط الفروق.

والجدول رقم (13 – 3) يبين لنا الخطوة الأولى المتعلقة بإجراء اختبار للعينات غير المستقلة، وذلك من خلال حساب الفروق في كمية مشاهدة الإذاعة المرئية لكل زوج.

جدول (13 ـ 3)

الأسرة	الفرق في دقائق المشاهدة (D)
1	22 = 23 – 45
2	31 = 25 – 56
3	30 = 43 – 73
4	27 = 26 – 53
5	6 = 21 – 27
6	5 = 29 – 34
7	44 = 32 – 76
8	−2 = 23 – 21
9	29 = 25 – 54
10	22 = 21 – 43
متوسط الفرق 21.4 = $\overline{X}D = \frac{\Sigma D}{N}$	

يمكننا ملاحظة أن متوسط الفرق مساوٍ للفرق بين المتوسطات، (وهو دائماً كذلك). أن السؤال الذي يمكن طرحه في هذا السياق هو، لماذا نقوم بهذا الإجراء البديل لحساب الفرق بين متوسطين؟. ومع أن متوسط الفرق سوف يكون دائماً مساوياً للفرق بين الوسطين، إلا أن التباين لن يكون الشيء نفسه، فالتباين حول متوسط الفرق يكون أصغر بكثير إذا قورن بالتباين حول الفرق بين الوسطين. والسبب في ذلك أنه من الممكن أن نعجز في رفض الفرق إذا ما تمت معاملته كفرق بين وسطين، عندما نكون قادرين على رفضه عندما نتعامل معه كمتوسط فرق.

وبالنظر إلى الجدول أعلاه فإننا نجد أن هناك في المتوسط فرقاً لكل الأزواج التي تشكل العينات. دعنا نفترض أنه في المجتمع ككل لا يوجد فرق في كمية المشاهدة للإذاعة المرئية بين الآباء وأبنائهم. وبالتالي يصاغ الفرض الصفري بالطريقة التالية:

$$H_O : \mu_D = 0$$

وعند إجراء عملية المعاينة من هذا المجتمع، فإننا قد نجد أحياناً أن أباً يشاهد الإذاعة المرئية أكثر مما يشاهدها ابنه، ولكن في أحيان أخرى قد نجد طفلاً يشاهد الإذاعة المرئية أكثر قليلاً عن عما يشاهده أبوه. ولكن إذا ما كان الفرض الصفري صحيحاً فإنه لا يوجد فرق في المتوسط فالفروق الموجبة سوف تلغي الفروق السالبة. بمعنى آخر، أنه من غير المعقول أن نتوقع أن التباين العشوائي قد يحدث أحياناً نتيجة لزيادة قليلة في الأسر التي يشاهد فيها الأب الإذاعة المرئية أقل من الطفل المناظر. أو العكس بالعكس، لذلك فإن فرق المتوسط بين العينات لن يكون صفراً (0)، فكلما كبر الفرق بين نتيجة العينة والنتيجة المتوقعة وهي صفر فرق المتوسط، مع ذلك، فإن احتمالاً ضئيلاً أن هذا سيكون راجعاً لتباين عشوائي، ولكن الأكثر احتمالاً أن هذا يعكس مبطن الفرق بين الآباء وأبنائهم.

في هذا المثال، فإن متوسط الفرق هو 21.4 دقيقة. فالسؤال المطروح في هذا الشأن هو هل هذا الفرق بين العينات يقودنا إلى رفض الفرض الذي مفاده لا يوجد فرق بين هذين المجتمعين؟

إن المعادلة المتعلقة بحساب اختبار t لفرق المتوسطات تكون:

$$t = \frac{\overline{X}_D}{S_D / \sqrt{N}}$$

حيث إن:

$$S_D = \sqrt{\frac{\Sigma D^2 - \frac{(\Sigma D)^2}{N}}{N-1}}$$

لاحظ أن (N) تشير إلى عدد الأزواج Number of Pairs وليس للعدد الكلي للحالات. في هذا المثال N = 10، بالرغم من أنه لدينا العدد الكلي الذي يصل إلى عشرين حالة، عشر حالات من الآباء، وعشر حالات من الأطفال.

إن درجة العينة لهذا المثال ستكون:

$$S_D = \sqrt{\frac{\Sigma D^2 - \frac{(\Sigma D)^2}{N}}{N-1}} = \sqrt{\frac{6410 - \frac{45.796}{10}}{10-1}}$$

$$= 14.2$$

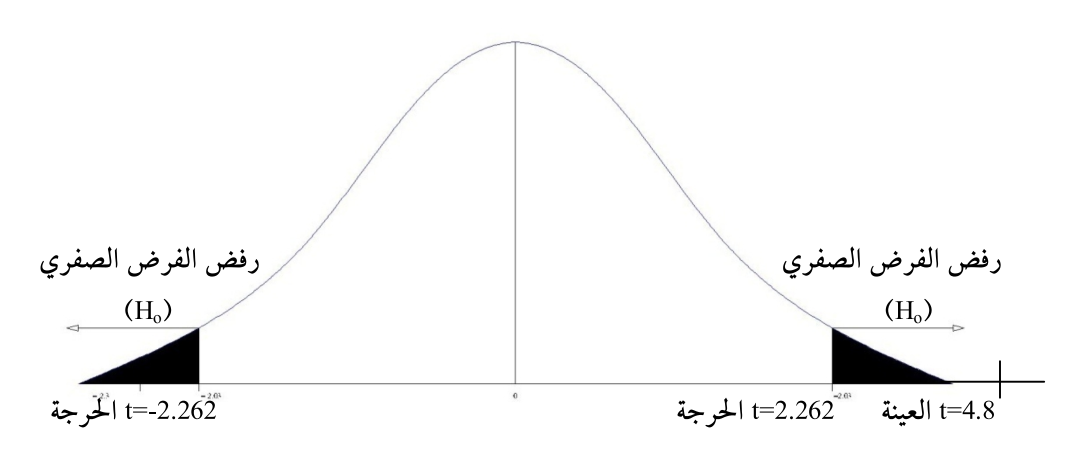
$$\text{العينة } t = \frac{\Sigma \overline{X}_D}{S_D/\sqrt{N}} = \frac{21.4}{14.2/\sqrt{10}}$$

$$= 4.8$$

في هذا المثال لدينا 9 درجات حرية (10 أزواج – 1) على مستوى دلالة (α) 0.05، فالقيمة الحرجة لـِ t تساوي:

$$\text{الحرجة } t = \pm 2.262\ (\alpha=0.05,\ df=9)$$

وإذا وضعنا هذه القيم في شكل رسم بياني (انظر الشكل رقم 13 – 6) فإننا نؤكد بأن الفرض الصفري قد تم رفضه. ومن هنا يمكننا القول، بأن هناك فرقاً دالاً في كمية مشاهدة الإذاعة المرئية بين الآباء والأبناء [11].



شكل رقم (13 ـ 6) الدرجة الحرجة ودرجة العينة

اختبار t لعينتين غير مستقلتين Paired-Samples t- test باستخدام SPSS :

الإجراء:

1- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Paired-Samples t- test ⇐ Copare means ⇐ Analyze

2- انقر على Minutes of TV Watched وبعد ذلك انقر على Parent

3- انقر على ◄ للصق المتغيرات في القائمة المحددة لِـ Paired Variables

4- انقر على OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

T. Test

Paired Samples Statistics

Pair 1 Minutes of TV Watched	Mean	N	Std. deviation	Std. Error Mean
Parent Minutes of TV Watched	26.80	10	6.85	2.10
Child	48.20	10	18.05	5.71

Paired Samples Correlation

Pair 1 Minutes of TV Watched	N	Correlation	Sig.
Parent Minutes of TV Watched Child	10	.899	.024

Paired Samples test

Pair 1 Minutes of TV Watched	Paired differences					t	df	Sig (2-tailed)
	Mean	Std. Deviation	Std Error Mean	confidence %95 Interval of the difference				
				Lower	Upper			
Parent Minutes of TV Watched Child	−21.40	14.22	4.50	−31.57	−11.23	4.758	9	.001

المصدر: George Argyrous, Statistics for Social & Health Reoearch, op.cit. , P. 439

شكل (13 ـ 7) مخرجات SPSS لاختبار t لعينتين ثنائيتين

تفسير مخرجات لاختبار t لعينتين ثنائيين:

من خلال هذه المخرجات نلاحظ أن هناك فرقاً ذا دلالة في كمية مشاهدة الإذاعة المرئية بين الآباء والأبناء والأطفال، حيث بيَّن مربع Paired Samples Statistics أن درجات الوسط الحسابي عمود Mean لكل مجموعة (الآباء والأبناء) فقد وصلت إلى 26.80 الوسط الحسابي للآباء، و 48.20، الوسط الحسابي للأبناء؛ وبالتالي يمكننا القول بأن هناك زيادةً ذات دلالة في كمية مشاهدة الإذاعة المرئية بين الآباء والأبناء. وبالتالي تحصلنا على قيمة t مساوية لـِ4.758 بدرجة حرية 9، وعليه نرفض الفرض الصفري الذي مفاده لا يوجد فرق.

كذلك تمدنا هذه المخرجات مربع Paired Samples test بـِ 95% فترة ثقة لتقدير الفرق. فالحد الأعلى هو 11.23-، بينما الحد الأدنى وصل إلى 31.57-. ويمكن للباحث استخدام هذه المعلومات لإجراء الاختبار. ولما كانت الفترة لا تحتوي قيمة صفر (0)، فإنه يمكننا القول، بأن الفرق في المجتمع فيما يتعلق بكمية مشاهدة الإذاعة المرئية للآباء وأطفالهم ليس صفراً.

ولمعرفة حجم التأثير لاختبار t لعينتين ثنائيتين، يمكننا حساب eta^2 من خلال إحدى المعادلتين:

$$eta^2 = \frac{t^2}{t^2 + N - 1} \qquad (1)$$

أو

$$r^2 = \frac{t^2}{t^2 + df} \qquad (2)$$

من خلال المعادلة رقم (1):

$$= \frac{4.758}{4.758^2 + 10 - 1}$$
$$= \frac{22.638}{22.638 + 9}$$
$$= \frac{22.638}{31.638}$$
$$= 0.715$$

من خلال هذه القيمة المتوصل إليها وهي ($eta^2 = .715$) يمكننا القول، بأن هناك تأثيراً كبيراً مع وجود فرق جوهري في درجات اختبار كمية مشاهدة الإذاعة المرئية للآباء والأطفال.

مثال إضافي لزيادة التوضيح[11]:

يرغب مدرس في معرفة تأثير وسيلة جديدة في التعليم، على قدرة التلاميذ في إكمال مادة الرياضيات الأساسية. فقد اختار هذا المدرس خمسة من الطلاب، وطلب منهم إكمال اختبار مادة الرياضيات الأساسية. وبعد ذلك أدخل الوسيلة الجديدة في التعليم. وبعد مضي شهرُ كاملُ عاد المدرس فاختار نفس الطلاب الخمسة، وطلب منهم إكمال نفس الاختيار السابق؛ وقد جاءت النتيجة كالتالي:

جدول (13 ـ 3) نتيجة اختبار مادة الرياضيات الأساسية

التلاميذ	وقت إنهاء الاختبار (قبلي)	وقت إنهاء الاختبار (بعدي)
أحمد	7.3	6.8
إبراهيم	8.5	7.9
جمعة	6.4	6.0
عادل	9.0	8.4
هاني	6.9	6.5
المتوسط (Mean)	$\overline{X} = 7.62$	$\overline{X} = 7.12$

لقد تعامل هذا المدرس أساساً مع هذين الاختبارين كعينتين مستقلتين. إن متوسط الوقت للاختبار القبلي هو 7.62 دقيقة. في حين أن متوسط الوقت للاختبار البعدي قد وصل إلى 7.12 دقيقة. وباستخدامنا لاختبار t للعينة يقدر بـِ 0.75، وهي درجة غير دالة على مستوى 0.05. وبالرغم من هذه النتيجة فقد رأى هذا المدرس أن هذه النتيجة تبدو نتيجة واعدة. فاختبار الاستدلال الذي تم إجراؤه غير قادر على رفض احتمالية أن هذا

التحسن قد جاء نتيجة للصدفة. وعليه فقد قرر هذا المدرس أن يتخلى عن هذه الطريقة الجديدة في التدريس.

ولحسن الحظ، أن أحد زملاء هذا المدرس يعرف أكثر قليلاً حول الإحصاء، وقد أدرك ذلك بما أن نفس الطلاب يشكلون كل عينة فإن إجراء اختبار عينات غير مستقلة أصبح أمراً ضرورياً لتصميم البحث. وبالتالي قد تم التعامل من خلال البيانات بالنتائج التالية:

الخطوة الأولى: صياغة الفرض الصفري والفرض البديل:

$$H_0 : \mu_D = 0$$

$$H_i : \mu_D \neq 0$$

الخطوة الثانية: اختيار اختبار الدلالة:

في هذا المثال، نحن نقارن بين عينتين غير مستقلتين فيما يتعلق بفروق المتوسط، وعليه، فإننا نستخدم t لعينتين غير مستقلتين لفروق المتوسط.

الخطوة الثالثة: حساب درجة العينة:

ولحساب فرق المتوسط بين العينتين وربط ذلك بدرجة t يمكننا توليد الجدول التالي:

التلاميذ	وقت إنهاء الاختبار (قبلي)	وقت إنهاء الاختبار (بعدي)	الفرق D	D^2
أحمد	7.3	6.8	0.5	0.25
إبراهيم	8.5	7.9	0.6	0.36
جمعة	6.4	6.0	0.4	0.16
عادل	9.0	8.4	0.6	0.36
هاني	6.9	6.5	0.4	0.16
المجموع Sum		$\sum D^2 = 1.29 \sum D = 2.5$		
المتوسط Mean		$\overline{X}D = 0.5$		

وبتعويض هذه البيانات لمعادلة الخطأ المعياري Standard error وبعد ذلك لمعادلة t. نحصل على النتائج التالية:

$$S_D = \sqrt{\frac{\Sigma D^2 - \frac{(\Sigma D)^2}{N}}{N-1}} = \sqrt{1.29 - \frac{(2.5)^2}{5-1}}$$

$$= 0.1$$

$$t = \frac{\overline{X}D}{S_D/\sqrt{N}} = \frac{0.5}{0.1/\sqrt{5}}$$

$$= 11.8$$

الخطوة الرابعة: إيجاد الدرجة الحرجة والمنطقة الحرجة على مستوى دلالة 0.05 بدرجات حرية 4، وأن القيمة الحرجة لـِ t تساوي (2.776).

الخطوة الخامسة: اتخاذ القرار:

عندما تم حساب درجة t على أساس أن العينتين غير مستقلتين بدلاً من أنهما مستقلتين، فالنتيجة المتحصل عليها، نتيجة ذات دلالة بشكل واضح على مستوى 0.01. ومن هنا بإمكان هذا المدرس أن يرفض الفرضية التي مفادها أن التحسن جاء فقط من خلال الفرضية العشوائية.

أسئلة للمراجعة:

1- ما هي الافتراضات التي ينبغي أن تكون حول توزيعات المجتمعات قبل إجراء اختبار t للعينات المستقلة؟

2- النتائج التالية: اختبر فرق الدلالة مستخدماً اختباراً ثنائيَّ الجانبين بألفا α تساوي 0.05 (مفترضين تباينات متساوية Assuming Equal Population Variance

حجم العينة	الانحراف المعياري	المتوسط الحسابي		
35	14.2	72	عينة 1	a)
50	11	76.1	عينة 2	
100	0.9	2.4	عينة 1	b)
100	0.9	2.8	عينة 2	
120	80	450	عينة 1	C)
100	77	475	عينة 2	

3- يرغب باحث في معرفة تأثير مكان الإقامة على العمر الذي يبدأ فيه الناس بالتريض. فقد قسم الباحث العينة العشوائية للمبحوثين إلى فئتين: فئة الريفيين (91) فرداً، وفئة الحضريين (107) فرداً. وقد وجد أن السكان الريفيين قد بدأوا بالتدخين عند متوسط عمري يقدر بـِ 15.75 سنة، بانحراف معياري 2.3 سنة. في حين أن السكان الحضريين قد بدأوا التدخين عند متوسط عمري 14.63 سنة، بانحراف معياري 4.1 سنة. السؤال الذي يمكن طرحه هو: هل هناك فرق ذو دلالة بين هاتين الفئتين؟ (استخدم تقدير التباين المجمع).

4- المؤسسة العامة للمياه والصرف الصحي، ترغب في تقييم فعالية الإعلان الإرشادي حول ترشيد استهلاك المياه في مدينة بنغازي لغرض التقليل من استهلاك المياه غير المرشد. قبل الإعلان اختارت المؤسسة عينة عشوائية من (100) أسرة من أنحاء مختلفة من المدينة، وتم تسجيل استخدامات الأسر من المياه في الفترة الصباحية بمتوسط يصل إلى 87 لتراً، بانحراف معياري 15 لتر. وقد تم سحب (100) أسرة

عشوائياً بعد الإعلان، حيث وصل معدل الاستخدام للمياه 74 لتراً. هل يوجد فرق ذو دلالة؟ ما هي النتيجة التي يمكننا التوصل إليها حول الإعلان الإرشادي؟ هل يمكننا إجراء اختبار أحادي الجانب أم ثنائي الجانب؟ ما هي العوامل التي يجب أن نأخذها في الاعتبار عند اختيار الاختبار المناسب؟

5- أ) ما هو فرق المتوسط للأزواج العشرة المشاهدة:

زوج	المشاهدة (1)	المشاهدة (2)
1	12	15
2	10	13
3	8	13
4	14	14
5	12	18
6	15	13
7	14	18
8	9	9
9	18	11
10	13	14

ب) ما هو الخطأ المعياري (S_D).

ج) إجراء اختبار t لعينتين غير مستقلتين بمستوى دلالة $\alpha = 0.05$.

6- اختبر الفرضيات التالية مستخدماً البيانات التالية:

H_O	H_i	فرق المتوسط	SD	N	α
a) $\mu D = 0$	$\mu D \neq 0$	2.3	1.4	20	0.10
b) $\mu D = 0$	$\mu D < 0$	-3.2	20	41	0.05

7- شركة تريد أن تستقصي ما إذا كان التغير في منظومة العمل يمكن أن يكون دالاً في التحسين في مستويات الإنتاج. فقد اختارت هذه الشركة عشرة أماكن للعمل وتم قياس مستويات الإنتاج فيما يتعلق بعدد الوحدات المنتجة في الساعة. وبعد ذلك أدخلت هذه الشركة برنامجاً لهذه الأماكن الإنتاجية مانحة المنتجين حرية التصرف في شروط وبناء الوظيفة، وتم قياس مستويات الإنتاج بعد 6 أشهر. وقد جاءت النتائج كالتالي:

مكان العمل	الإنتاجية قبل التغيير	الإنتاجية بعد التغيير
1	120	165
2	121	154
3	145	120
4	112	155
5	145	164
6	130	132
7	134	154
8	126	162
9	137	130
10	128	142

- السؤال المطروح هو: هل إدخال البرنامج الجديد قد أدى بشكل دالٍ إلى تحسن مستويات الإنتاج (لاحظ صياغة الفرض البديل)؟

الهوامش والمصادر:

أولا الهوامش:

1- George Argyrous, Statistics for Social & Health Research, with a Guide to SPSS , SAGE Publications, London, 2001, P. 341.

2- Ibid. , P. 344.

3- Ibid. , P. 347.

4- Ibid. , PP. 349 - 350.

5- جولي بالانت (ت) خالد العامري، التحليل الإحصائي باستخدام برنامج SPSS، دار الفاروق للنشر والتوزيع، ط 2، مصر، 2009، ص 225.

6- Ibid. , P. 233.

7- Ibid. , P. 347.

8- Ibid. , PP. 351 - 352.

9- عبد الله عامر الهمالي، التقنيات الإحصائية ومناهج البحث، منشورات جامعة قاريونس، 2008 م.

10- George Argyrous , op.cit. , P. 434.

11- Ibid. , PP. 434 - 437.

12- Ibid. , PP. 440 - 442.

ثانيا: المصادر:

1- Frederick J. Gravetter and Larry B. Wallnau , Statistics for The Behavirol Sciences ,8th ed , Wadsworth Cengage Learning , USA, 2010.

2- George Argyrous, Statistics for Social & Health Research, with a Guide to SPSS , SAGE Publications, London, 2001.

3- جولي بالانت (ت) خالد العامري، التحليل الإحصائي باستخدام برنامج SPSS، دار الفاروق للنشر والتوزيع، ط 2، مصر، 2009.

4- عبد الله عامر الهمالي، التقنيات الإحصائية ومناهج البحث، منشورات جامعة قاريونس، 2008 م.

الفصل الرابع عشر

فروق الدلالة : اختبار F لأكثر من وسطين متساويين :

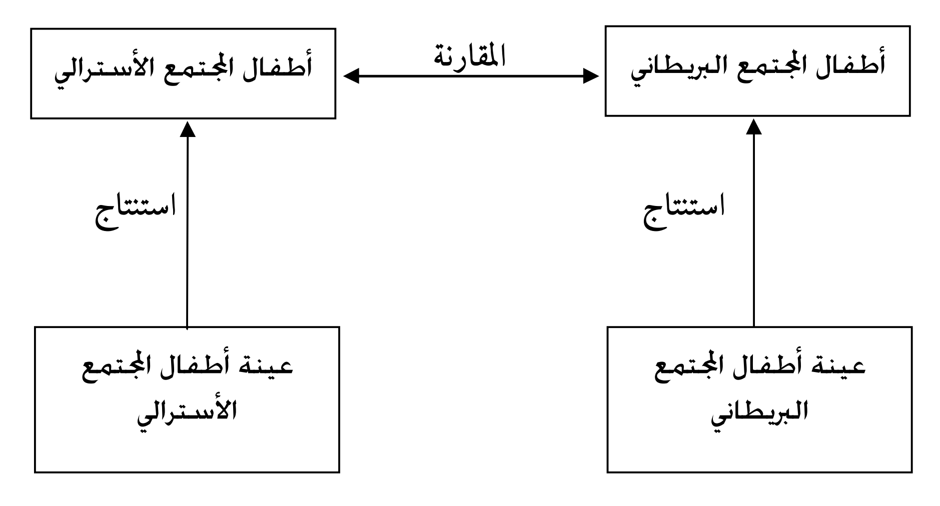
"تحليل التباين"

اختبار الفرض لأكثر من عينتين : الفكرة العامة :

تناولنا في الجزء الأول من الفصل السابق اختبار t لعينتين مستقلتين واختبرنا الافتراض بأن العينات قد تم سحبها من مجتمعات متوسطاتها متساوية:

$$Ho : \mu_1 = \mu_2$$

ولقد تعاملنا مع مثال حيث كانت العينة مكونة من عشرين طفلاً من المجتمع الأسترالي وعشرين طفلاً من المجتمع البريطاني. وقد تَمَّ طرح سؤال على كل طفل في هاتين العينتين يطلب فيه من كل طفل أن يحدد عدد الدقائق التي يشاهد فيها الإذاعة المرئية في الليلة الواحدة. وقد تمت مقارنة العينتين من أجل اختبار الفرض الصفري Ho، الذي مفاده أنه لا يوجد فرق في متوسط كمية مشاهدة الإذاعة المرئية (التلفزيون) بين هؤلاء الأطفال في كلا المجتمعين (انظر الشكل التالي):

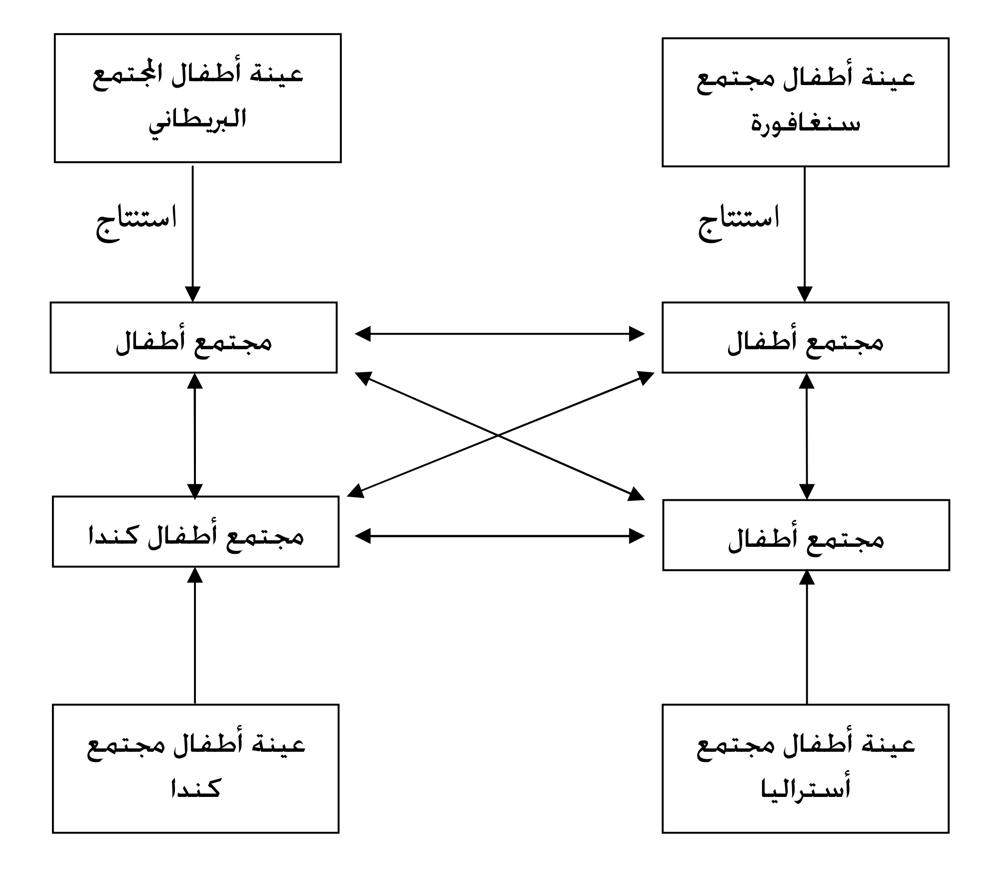


شكل (14 ـ 1) اختبار فرض لحالة عينتين

ويطلق على هذه العملية مسألة العينتين لأننا نتعامل مع عينتين من أجل الاستدلال حول كل مجتمع. إلا أننا في بعض الأحيان نتعامل مع مسألة أوسع بقليل. وبدلاً من مجرد المقارنة بين مجتمعين، فإننا قد نرغب في مقارنة كمية متوسط ما تمت مشاهدته في الإذاعة المرئية من قِبل الأطفال في عدة مجتمعات. على سبيل المثال، يمكن أن يكون لدينا عينة مؤلفة من عشرين طفلاً (20) تمثل المجتمعات التالية: أستراليا، بريطانيا، كندا وسنغافورة، ونود معرفة ما إذا كانت المتوسطات الحسابية لكل هذه المجتمعات متساوية:

$$Ho : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

ويطلق على هذه المسألة عينات K المستقلة K independent samples، حيث إن K تمثل أي عدد يزيد عن اثنين. وفي هذا المثال فإن K تساوي أربعة، ويمكننا توضيح هذا المثال بالشكل التالي:



المصدر: George Argyrous, Statistics for Social & Health Research,op. Cit , p. 356.

شكل (14 ـ 2) اختبار فرض لأكثر من عينتين

إن أحد الطرق للقيام بعملية المقارنة هي إجراء اختبار لكل عينتين متحدتين معاً مع العينات الأربعة بحيث يكون الحد الأعلى لهذه التركيبة هو 6، كما تم بيانه في الشكل أعلاه، فالأسهم تربط هذه المجتمعات بعضها ببعض:

المجتمع الأسترالي مع المجتمع السنغافوري

المجتمع الأسترالي مع المجتمع الكندي

المجتمع الأسترالي مع المجتمع البريطاني

المجتمع السنغافوري مع المجتمع الكندي

المجتمع السنغافوري مع المجتمع البريطاني

المجتمع الكندي مع المجتمع البريطاني

إذاً، يمكننا إجراء ستة اختبارات منفصلة لتقييم ما إذا كان هناك أي فروق دالة، تجدر الإشارة إلى أنه عندما نتعامل مع أكثر من عينتين، فإننا بذلك نستطيع اختبار متوسطات متساوية كلها مرة واحدة مستخدمين تحليل التباين لاختبار F (أنوفا ANOVA). والسبب وراء تفضيل استخدام تحليل التباين الأحادي على اختبارات t المتعددة يكمن في الوقوع في الخطأ من النوع الأول Type I error لسلسة من اختبارات t ستكون أكبر من مستوى الدلالة المقررة لكل اختبار متعلق بـِ t. وعليه، إذا كان مستوى ألفا (α) لكل اختبار على حدة لـِ t هو 0.05، فإن فرصة الوقوع في الخطأ من النوع الأول لكل اختبارات t التي يمكن إجراؤها لعدد معين من العينات سوف يكون أكبر من 0.05 على الجانب الآخر، فإن اختبار أنوفا ANOVA، ينص على أن مستوى ألفا مساوٍ لخطر الوقوع في الخطأ من النوع الأول[1].

إن إجراء أنوفا لاختبارات الفرض الصفري الذي يشير إلى أن العينات سحبت من مجتمعات ذات متوسطات متساوية. فإذا كان الفرض الصفري صحيحاً، فإن العينات المسحوبة من مثل هذه المجتمعات سوف تكون قيمة متوسطاتها تقريباً متساوية. ففي مثال الأطفال ومعدل مشاهدة الإذاعة المرئية (التلفزيون)، فإن كل العينات لديها تقريباً متوسطات متساوية. وبطبيعة الحال، فإننا لا نتوقع أن تكون متوسطات العينة متساوية، حتى ولو كانت متوسطات المجتمع متساوية، حيث إن التباين العشوائي Random Variation سوف يؤثر على عملية المعاينة Sampling Process. إن السؤال الذي نطرحه هو ما إذا كانت الفروق بين العينات فروقاً متسقةً مع فرضية التساوي Assumption of Equality بين المجتمعات. إذا أخذنا بعين الاعتبار نتائج العينة الافتراضية للمجموعات الأربعة من الأطفال كما هو موضح في الجدول التالي:

جدول (14ـ1)

متوسط كمية مشاهدة الإذاعة المرئية في الليلة الواحدة (بالدقائق)

مشاهدة الإذاعة المرئية في الليلة الواحدة	المجتمع			
	كندا	أستراليا	بريطانيا	سنغافورة
المتوسط	127	166	187	203
الانحراف المعياري	27	29	30	26

المصدر: George Argyrous, Statistics for Social & Health Research,op. Cit , p. 357.

فإنه يمكننا ملاحظة أن هناك تبايناً واضحاً بين متوسطات هذه العينات الأربعة. وحقيقة الأمر، إذا قارنا القيم العليا بالقيم الدنيا لكل من كندا وسنغافورة، فإننا سنجد أن هناك فرقاً كبيراً في كمية متوسط مشاهدة الإذاعة المرئية. كما يمكن أيضاً ملاحظة (الصف المتعلق بالانحراف المعياري لكل عينة من العينات الأربعة) أن النتائج داخل عينة في كل مجتمع تتجمع معاً كما أشير إليها بانحرافات معيارية صغيرة نسبة إلى المتوسطات. بمعنى آخر، أنه توجد فروق واضحة من بلد إلى بلد آخر، ولكنها تتشابه داخل كل بلد. وفي ضوء هذه الإحصاءات الوصفية، فإنه بإمكاننا أن نبدأ بطرح تساؤل حول الفرض الذي مفاده أن المجتمعات التي من خلالها تم سحب هذه العينات هي مجتمعات لديها نفس المتوسط.

إن هذا المنطق يشبه تماماً ذلك المنطق المستخدم في تحليل التباين الذي يقوم على أساس مقارنة كمية التباين بين العينات مع كمية التباين داخل كل عينة من العينات. إذاً، بالرغم من أننا نرغب في معرفة الفرق بين المتوسطات، ففي واقع الأمر، فإن تحليل التباين يتعامل مع التباين الذي هو تربيع الانحراف المعياري.

تحليل التباين الأحادي: (اختبار F):

يمكننا الآن استخدام المفاهيم العامة لتحديد ما إذا كان هناك فرق دال بين الأطفال في مجتمعات مختلفة فيما يتعلق بكمية متوسط مشاهدتهم للإذاعة المرئية. ولحساب إحصاء الاختبار المناسب، فإننا نحتاج إلى بيان بعض المفاهيم الأساسية. أولها: كمية التباين الكلي

للدرجات المتعلقة بكل الحالات الثمانين (80) التي تم معاينتها. وقد تم قياس كمية التباين الكلي من خلال مفهوم يطلق عليه: **مجموع التربيعات الكلية** TSS. ويتم حساب التربيعات الكلية وفقاً للمعادلة التالية:

$$TSS = \sum (X - \overline{X})^2$$

إن قيمة مجموع التربيعات الكلية Tss يمكن أن تُقَسَّمَ إلى عنصرين، العنصر الأول هو كمية التباين داخل كل عينة، ويطلق عليه مجموع التربيعات داخل المجموعات (SSW). والعنصر الثاني هو كمية التباين بين كل عينة، ويطلق عليه مجموع التربيعات بين المجموعات (SSB):

$$\text{TSS} = \text{SSB} + \text{SSW}$$

إن كل عنصر من هذه العناصر المتعلقة بمجموع التربيعات الكلية TSS يمكن حسابها بالطريقة التالية:

$$SSW = \sum (X - \overline{X}_S)^2$$

$$SSB = \sum n_S (\overline{X}_S - \overline{X})^2$$

حيث إن: X_s متوسط العينة المعطاة.

: n_s عدد الحالات في العينة المعطاة.

تجدر الإشارة إلى أن هذه المعادلات تذكرنا بالمعادلة المتعلقة بالانحراف المعياري حيث إن هذه المعادلة تتضمن نفس الأساس بأن عملية التباين مرتبطة بكمية الفرق بين الدرجات الفردية والمتوسط، أما فيما يتعلق بمعادلة الانحراف المعياري، فإن هذه المعادلات المعَرَّفة Definitional Formulas، من الصعوبة بمكان العمل بها. وبشكل خاص، إذا أردنا أن نحسب المجموع الكلي للتربيعات TSS، فإنه من السهولة بمكان العمل بالمعادلة التالية:

$$TSS = \sum x^2 - n\bar{x}^2$$

ولكي أن نتحصل على مجموع التربيعات الكلية، فإننا بحاجة فقط لحساب إما SSW أو SSB وحينها نستخدم المعادلة التالية: TSS=SSB+SSW لحساب الأخرى. بمعنى آخر، إذا تم حساب قيمة TSS و SSB فإننا نستعيض بهذه الحسابات للمعادلة التالية لكي نتوصل إلى قيمة SSW:

$$SWW = TSS - SSW$$

ولكي نعرف كيف تم ذلك، دعنا نعيد العمل من خلال المثال السابق للعينات الأربعة المتعلقة بالعشرين طفلاً. ولإجراء هذه العمليات الحسابية بشكل جيد، يتطلب الأمر بناء جدول يحتوي على قائمة بيانات (انظر الجدول 2).

من خلال هذا الجدول سنجد أن درجة كل حالة من الحالات قد دونت، وأن العينات قد وضعت في أعمدة منفصلة. ومن خلال هذه المعلومات يمكننا حساب المتوسط الحسابي لكل عينة من هذه العينات، إضافة إلى المتوسط الحسابي لكل العينات موحدة.

كندا $\bar{X} = \frac{2546}{20} = 127.3$ دقيقة

أستراليا $\bar{X} = \frac{3317}{20} = 165.85$ دقيقة

بريطانيا $\bar{X} = \frac{3735}{20} = 186.75$ دقيقة

سنغافورة $\bar{X} = \frac{4063}{20} = 203.15$ دقيقة

جدول (14ـ2) لحساب تحليل التباين (ANOVA)

كندا		أستراليا		بريطانيا		سنغافورة	
X	X^2	X	X^2	X	X^2	X	X^2
89	7921	102	10,404	124	15,376	156	24,336
92	8464	120	14,400	135	18,225	165	27,225
95	9025	132	17,424	156	24,336	174	30,276
105	11,025	134	17,956	165	27,225	179	32,041
106	11,236	145	21,025	167	27,889	180	32,400
108	11,664	149	22,201	172	29,584	184	33,856
110	12,100	156	24,336	178	31,684	189	35,721
113	12,769	162	26,244	182	33,124	189	35,721
116	13,456	165	27,225	184	33,856	196	38,416
125	15,625	165	27,225	185	34,225	203	41,209
128	16,384	165	27,225	186	34,596	204	41,616
135	18,225	174	30,276	187	34,969	207	42,849
138	19,044	178	32,041	189	35,721	210	44,100
139	19,321	180	32,400	198	39,204	218	47,524
140	19,600	187	34,969	209	43,681	221	48,841
146	21,316	189	35,721	212	44,944	228	51,984
146	21,316	196	38,416	218	47,524	231	53,361
154	23,716	201	40,401	223	49,729	238	56,644
167	27,889	206	42,436	225	50,625	241	58,081
194	37,636	210	44,100	240	57,600	250	62,500
$\sum X = 2546$		$\sum X = 3317$		$\sum X = 3735$		$\sum X = 4063$	
$\sum X^2$ =337,732		$\sum X^2$ =566,425		$\sum X^2$ =714,117		$\sum X^2$ =838,701	

المصدر: George Argyrous, Statistics for Social & Health Research, with a Guide to SPSS , SAGE Publications, London , 2001 , p. 361.

$$\overline{X} = \frac{(2546+3317+3735+4063)}{80} = 170.8$$ دقيقة

وباستخدام هذه المعلومات يمكننا حساب TSS و SSB.

$$TSS = \sum X^2 - n\overline{x}^2$$

$$= (337{,}732 + 566{,}425 + 714{,}117 + 838{,}701) - (80)\,(170.8)^2$$

$$= 124{,}189$$

$$SSB = \sum n_S(\overline{x}_S - \overline{x})^2$$

$$= 20\,(127.3 - 170.8)^2 + 20\,(165.85 - 170.8)^2$$

$$+ 20\,(186.75 - 170.8)^2 + 20\,(203.15 - 170.8)^2$$

$$= 64{,}353$$

$$SSW = TSS - SSB = 124{,}189 - 64{,}353$$
$$= 59{,}836$$

إن إحصاء الاختبار الفعلي الذي يمكن استخدامه لتحديد ما إذا كان هناك فرق دال في تقدير التباين بين العينات وداخل العينات، نستخدم نسبة F (F.ratio). وتعني F.ratio نسبة تباين الاثنين: SSB و SSW وكل واحدة من هذه تصحح بدرجات حرية ملائمة:

العينة $F = \dfrac{SSB/K-1}{SSW/n-K}$

حيث إن: K تشير إلى عدد العينات.

وبتعويض الأرقام ذات الصلة لهذه المعادلة نحصل على:

العينة $F = \dfrac{SSB/K-1}{SSW/n-K}$

$$= \frac{64,353/4-1}{58,836/80-4}$$

$$= 27.25$$

وكما رأينا عند التعامل مع اختبارات Z، و T و X^2 لاحقاً، فإننا بحاجة إلى مقارنة الدرجة المحسوبة بالدرجة الاحتمالية من أجل الوصول إلى قرار يمكن من خلاله أن نقبل أو نرفض الفرض الصفري Ho.

ولإيجاد القيمة الحرجة يمكننا إحالة القارئ إلى جدول توزيع القيم الحرجة لـِ F على مستوى دلالة 0.05. انظر الجدول (14 – 4).

جدول (14 ـ 3): تحليل التباين الأحادي
(دقائق مشاهدة الإذاعة المرئية في الليلة الواحدة)

مصادر التباين	درجة الحرية	مجموع المربعات	متوسط مجموع المربعات	اختبار (F)	درجة الدلالة 0.05 (α)
بين المجموعات	3	64353.438	21451.146	27.246	.000
داخل المجموعات	76	59835.050	787.303		
المجموع الكلي	79	124188.488			

وللحصول على القيم الحرجة لاختبار F ينبغي على الباحث أن يراعي ثلاثة عوامل أساسية:

1- درجات الحرية لتقدير التباين بين العينات. وللحصول على درجة الحرية: عدد العينات ناقص واحد. ويظهر هذا في البسط المتعلق بِـ F- ratio:

$$Df_b = k - 1 = 4-1$$

$$= 3$$

$$Df_w = n - k = 80 - 4$$

$$= 76$$

جدول (14 ـ 4) القيم الحرجة لتوزيعات F (ألفا 0.05 = α)

	درجات الحرية لتقدير التباين بين عينات K-1									
n-k	1	2	3	4	5	6	7	8	9	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.84	8.81	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.63
5	6.16	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.36
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	1.51
50	4.03	3.18	2.79	2.56	2.38	2.29	2.20	2.13	2.07	1.44
60	4.00	3.15	2.76	2.52	2.37	2.25	2.17	2.10	2.04	1.39
80	3.96	3.11	2.72	2.48	2.33	2.19	2.12	2.05	1.99	1.32
100	3.94	3.09	2.70	2.46	2.30	2.17	2.10	2.03	1.97	1.28
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.25
α	3.84	2.99	2.60	2.37	2.21	2.09	2.01	1.94	1.88	1.00

من خلال جدول توزيع القيم الحرجة لِـ F على مستوى دلالة 0.05 يمكننا أن نلاحظ أنه لا يوجد صف متعلق بدرجة الحرية داخل درجة تساوي 76. في الحقيقة أن مدى كل القيم قد قفزت بعد أول 30؛ ويرجع السبب في ذلك، إلى أن درجات الحرية لا تقل كثيراً عن زيادة مقدار في درجة الحرية بعد الثلاثين. وعندما لا تكون لدينا درجات حرية تظهر في الجدول، فإننا نشير إلى أقرب قيمة تظهر في الجدول من الرقم المرغوب فيه: في هذا المثال، إن أقرب قيمة دون 76 تظهر في الجدول هي عند الرقم 60. وعليه، فإن القيمة الحرجة لِـ F تكون في المثال هي 2.76. ومن هنا نرفض الفرض الصفري Ho الذي مفاده لا فرق باعتبار أن قيمة F المحسوبة أكبر من قيمة F المتوقعة:

نرفض الفرض الصفري عندما تكون (F المتوقعة > F العينة).

وعليه عند رفضنا للفرض الصفري فإنه على الأقل واحد من مجتمعات هؤلاء

الأطفال تختلف عن الأخريات. لاحظ التعبير الخاص للنتيجة: على "الأقل" واحد من هذه المجتمعات يختلف عن باقي المجتمعات الأخرى.

إن اختبار F في حد ذاته لا يخبرنا أي المجتمعات، وكم كمية الاختلاف. وبكل وضوح إذا كانت هناك فروق، إذاً على الأقل يجب أن تشمل المجموعات التي لديها عينة بمتوسطات عليا ودنيا في هذا المثال ستكون كندا وسنغافورة على التوالي.

وكما أوضحنا سابقاً برفضنا للفرض الصفري بعد إجراء اختبار F فإننا بذلك قد قررنا أن واحداً على الأقل من هذه المجتمعات لديه متوسط حسابي غير مساوٍ لمتوسطات المجتمعات الأخرى. وعليه فإن اختبار F في حد ذاته لا يخبرنا أي هذه المجتمعات يكون مختلفاً. ولكي نحدد أياً من المجتمعات يختلف عن الأخرى، فالباحث يمكنه أن يلجأ إلى استخدام Post - Hoc Comparisons المتوفرة كخيار في برنامج SPSS[(2)].

إجراء اختبار F باستخدام Spss:

1- من القائمة الموجودة في الجزء العلوي من الشاشة انقر فوق:

Analyze ⟸ **Compare Means** ⟸ **one Way ANOVA**

2- انقر فوق المتغير التابع **Minutes of TV Watched** في قائمة المتغير.

3- انقر على ◄ مشيراً إلى الصندوق أسفل **Dependent List**. هذه القائمة تقود إلى لصق **Minutes of TV Watched** في القائمة المحددة للمتغير التابع الذي يستخدم لمقارنة العينات.

4- انقر على **Country of Residence** في القائمة.

5- أنقر على ◄ مشيراً إلى الصندوق أسفل: **Factor** تقود هذه العملية إلى لصق **Country of Residence** في القائمة المحددة لمتغير **Factor** الذي يشكل العينات المستهدفة للمقارنة.

6- انقر على ok.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

ANOVA

ONE Way Minutes of TV Watched Per night

	Sum of Squares	dF	Mean Square	F	Sig.
Between groups	64353.438	3	21451.146	27.246	.000
Within groups	59835.050	76	787.303		
TOTAL	124188.488	79			

المصدر: George Argyrous, op. , p. 361.

شكل (14 ـ 3) مخرجات لتحليل التباين الأحادي

تفسير مخرجات SPSS لتحليل التباين:

بالنظر إلى مخرجات SPSS لتحليل التباين يمكننا أن نلاحظ في المربع أعلاه مجموع التربيعات بين المجموعات Sum of Squares Between Groups، ومجموع التربيعات داخل المجموعات Sum of Squares Within Groups، ومجموع التربيعات الكلية The Total Sum of Squares في العمود الأول في مربع ANOVA معاً مع درجة الحرية المناسبة (df) في العمود الثالث. ومن خلال هذه المخرجات نجد أن قيمة F تساوي 27.246 (العمود 5)، وأن درجة الدلالة عمود .Sig تساوي 000.، ولا تعني هذه الدلالة أن احتمالية الحصول على نسبة 27.25 لِـ F تساوي صفراً (0)، ذلك أن برنامج SPSSيقرب الاحتمالية إلى ثلاث درجات عشرية decimal Places3 ، وعليه، فإن هذه الاحتمالية تقرأ أقل من 5 في 10.000.

ولَمّا كانت القيمة في هذا المثال (Sig. = 0.000)، وهي أقل من 05. مما يدل على وجود فروق دالة بين المجموعات.

حجم التأثير:

ولحساب حجم التأثير الذي يعطي مؤشراً لحجم الفروق بين المجموعات (وليس مجرد تحديد ما إذا كان الفارق قد حدث على سبيل الصدفة أم لا) يستخدم إحصاء Eta^2

إيتا تربيع. ولما كانت مخرجات SPSS لا تقوم بحساب Eta^2 في اختبار F، فالباحث يمكنه حسابها من المخرجات المرتبطة باختبار F، وفقاً للمعادلة التالية:

$$Eta^2 = \frac{\text{مجموع التربيعات بين المجموعات}}{\text{المجموع الكلي للمربعات}}$$

$$Eta^2 = \frac{SSb}{TSS} = \frac{64353.438}{124188.488}$$
$$= 0.518$$

من هنا نجد أن حجم التأثير يبلغ 0.518 أو (52 %) وهو تأثير كبير نسبياً حسب إرشادات كوهين، مع وجود فرق دال في متوسط عدد الدقائق المشاهدة في الليلة الواحدة.

مثال لزيادة التوضيح[3]:

في دراسة قام بها أحد الباحثين على ثلاثة أطفال لمقارنتهم في القدرة على القراءة. وقد طلب من كل طفل قراءة 12 واجباً مع ضبط الأخطاء خلال قراءة كل واجب من هذه الواجبات الاثنى عشر، ويود الباحث معرفة هل هؤلاء الأطفال يختلفون في قدرات القراءة؟ ويوضح الجدول التالي النتائج المتحصل عليها:

جدول (14-4) عدد الأخطاء لكل طفل

رقم الواجب	زيد	عمرو	خالد
1	8	15	12
2	6	9	6
3	14	20	8
4	9	15	9
5	14	6	10
6	8	9	14
7	12	17	16
8	19	12	5
9	6	6	18
10	11	3	21
11	8	13	15
12	15	5	11

من خلال هذه البيانات، هل يمكننا القول بأن هناك فروقاً ذات دلالة بين هؤلاء الأطفال الثلاثة في قدرات القراءة؟

للإجابة على هذا السؤال يمكننا إجراء اختبار F وذلك باتباع الخطوات التالية:

الخطوة الأولى: صياغة الفرض الصفري والفرض البديل:

Ho: متوسط عدد الأخطاء التي ارْتُكِبَتْ من قِبل كل طفل من هؤلاء الأطفال هي متوسطات متساوية: $Ho: \mu_1 = \mu_2 = \mu_3$

Hi: متوسط عدد الأخطاء التي ارْتُكِبَتْ من قِبل كل طفل من هؤلاء الأطفال ليست كلها متوسطات متساوية: $Hi: \mu_1 \neq \mu_2 \neq \mu_3$

الخطوة الثانية: اختيار اختبار الدلالة:

إن السؤال البحثي الذي يرغب الباحث فيه هو متوسط عدد الأخطاء التي تم قياسها على مستوى المقياس ذي المسافات والنسبي. وعليه فالرغبة تنحصر في مقارنة المتوسطات وبيان ما إذا كانت هذه المتوسطات متساوية. ولما كانت لدى الباحث ثلاث عينات لذلك ستكون المقارنة عبر أكثر من عينتين. وأن الاختبار الملائم لهذه المسألة البحثية هو تحليل التباين (اختبار F لمتوسطات متساوية).

الخطوة الثالثة: حساب درجة العينة:

لإجراء تحليل التباين ANOVA، فإنه من المفيد أن يوجد جدول يحتوي على الحسابات المرغوبة (انظر الجدول 5)، ومن هذه المعلومات نقوم بحساب المتوسط الحسابي لكل عينة من العينات الثلاثة، وبعد ذلك تجمع المتوسطات لكل هذه العينات.

زيد $\overline{X} = \frac{130}{12} = 10.8$

عمرو $\overline{X} = \frac{140}{12} = 11.7$

خالد $\overline{X} = \frac{145}{12} = 12.1$

جدول (14 ـ 5) حساب تحليل التباين ANOVA

زيد		عمرو		خالد	
X	X^2	X	X^2	X	X^2
8	64	15	225	12	144
6	36	9	81	6	36
14	196	20	400	8	64
9	81	15	225	9	81
14	196	6	36	10	100
8	64	9	81	14	196
12	144	17	289	16	256
19	361	12	144	5	25
6	36	6	36	18	324
11	121	13	169	21	441
8	64	13	169	15	225
15	225	5	25	11	121
$\Sigma X = 130$	$\Sigma X^2 = 1588$	$\Sigma X = 140$	$\Sigma X^2 = 1880$	$\Sigma X = 145$	$\Sigma X^2 = 2013$

$$\overline{X} = \frac{130 + 140 + 145}{36} = 11.5$$

وتمثل هذه العمليات الحسابية الإحصاء الوصفي لبيانات العينة. وبوضوح بأنه يوجد فرق بين العينات فيما يتعلق بمتوسط عدد الأخطاء المرتكبة. هل هذه الأخطاء المرتكبة راجعة إلى تباين عشوائي عند عملية المعاينة من مجتمعات لا يوجد فرق فيما بينها؟

ولكي نقرر ذلك، يمكننا بادئ ذي بدء حساب TSS و SSB:

$$TSS = \Sigma X^2 - n\bar{x}^2 = (1588 + 1880 + 2013) - 36(11.5)^2$$

$$= 720$$

$$SSB = \Sigma ns(\bar{x}_S - \bar{x})^2$$

$$= 12\,(10.8 - 11.5)^2 + 12\,(11.7 - 11.5)^2 + 12(12.1 - 11.5)^2$$

$$= 10.7$$

$$SSW = TSS - SSB = 720 - 10.7$$

$$= 709.3$$

ومن خلال هذه البيانات يمكننا حساب قيمة F الإحصائية وذلك باستخدام اختبار الدلالة:

$$F_{\text{العينة}} = \frac{SSB/K-1}{SSW/n-K}$$

$$= \frac{10.7/3-1}{709.3/36-3}$$

$$= 0.25$$

الخطوة الرابعة: إيجاد الدرجة الحرجة والمنطقة الحرجة:

من خلال جدول توزيع F الحرجة، فإن K=2 و N=36

وعليه، فإن القيمة الحرجة لـِ F:

$$F_{\text{الحرجة}} = 3.32\ (\alpha = 0.05)$$

الخطوة الخامسة: اتخاذ القرار:

إذا أردنا رسم درجة العينة ودرجة القيمة الحرجة في الشكل رقم (14 – 3)، فإننا بذلك لا يكون في مقدورنا رفض الفرض الصفري.

حجم التأثير: ولمعرفة مؤشر حجم الفروق بين هذه المجموعات (ليس مجرد تحديد ما إذا كان الفارق قد حدث بالصدفة أم لا)، يمكننا استخدام إحصاء Eta^2 إيتا تربيع التي يمكن حسابها وفقاً للمعادلة التالية:

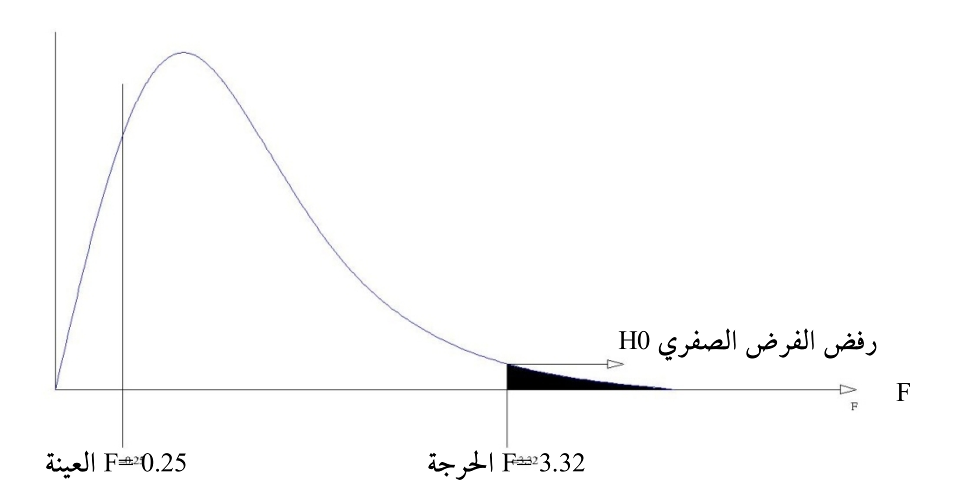
$$\text{إحصاء إيتاء تربيع} = \frac{\text{مجموع التربيعات بين المجموعات}}{\text{المجموع الكلي للمربعات}}$$

$$\text{Eta}^2 = \frac{ssb}{Tss}$$

$$= \frac{10.7}{720}$$

$$= 0.015$$

من هنا نجد أن حجم التأثير يبلغ 0.015 في المائة وهو تأثير ضئيل حيث يشير إلى أنه لا يوجد فرق جوهري في متوسط عدد الأخطاء لدى عينات هؤلاء الأطفال.



شكل (14-3) درجات العينة والحرجة

أسئلة للمراجعة:

1- لقد تمت المقارنة بين أربع جمعيات للرعاية في إطار متوسط عدد الحالات التي يقدمها موظفو هذه الجمعيات خلال شهر. ويهدف البحث للإجابة عن السؤال ما إذا كان الفرق بين هذه الجمعيات وفقاً لعبء العمل ذا دلالة.

أ- اشرح لماذا يستخدم تحليل التباين ANOVA لسبر غور هذه المسألة.

ب- بين الفرض الصفري لهذه المسألة لفظياً وجبرياً.

ج- من النتائج الافتراضية احسب قيمة F واتخذ القرار المناسب حول الفرض الصفري (Ho) بمستوى دلالة (α = 0.05).

التباين	مجموع التربيعات	درجة الحرية
بيــن الجمعيات	50	4
داخل الجمعيات	7210	110
المجموع الكلي	**7260**	**114**

2- مُحَاضِرٌ بالجامعة يستخدم طرق تدريس مختلفة في ثلاثة فصول دراسية منفصلة. وكان الهدف من وراء هذا الاستخدام تقييم الفعالية النسبية لهذه الطرق باختبار فرق الدلالة بين هذه الفصول. والدرجات النهائية تظهر في الجدول التالي:

الطريقة (أ)	الطريقة (ب)	الطريقة (ج)
21	28	19
19	28	17
21	23	20
24	27	23
25	31	20
20	38	17
27	34	20
19	32	21
23	29	22
25	28	21
26	30	23

المطلوب:

أ- احسب المتوسط الحسابي، والانحراف المعياري لكل عينة من العينات الثلاث. هل تستطيع أن تتوقع من هذه الإحصاءات الوصفية النتيجة المتحصل عليها من خلال إجراء أنوفا ANOVA على هذه البيانات.

ب- أَجْرِ تحليل ANOVA لتقييم هذه التوقعات.

3- البيانات التالية، بيانات افتراضية حول عينة مؤلفة من 20 طفلاً من مجتمع الولايات المتحدة الأمريكية تبين عدد الدقائق التي يشاهدها هؤلاء الأطفال في الليلة الواحدة:

195	184	165	162	168	196	217	190	212	232
204	205	217	210	230	197	180	192	190	198

المطلوب:

أ- كيف تُؤثر إضافة هذه العينة على اختبار تحليل التباين ANOVA على أطفال مجتمعات: أستراليا، بريطانيا، كندا، وسنغافورة، على عدد درجات الحرية؟

ب- ما النتيجة التي يمكنك التوصل إليها فيما يتعلق بكمية مشاهدة الأطفال للإذاعة المرئية من بلدان مختلفة؟

هوامش والمصادر:

أولا: الهوامش:

1- George Argyrous, Statistics for Social & Health Research, with a Guide to SPSS , SAGE Publications, London, 2001, P. 357.

2- Ibid. , PP. 364 - 365.

3- Ibid. , PP. 367 - 370.

ثانيا: المصادر:

1- George Argyrous, Statistics for Social & Health Research, with a Guide to SPSS , SAGE Publications, London, 2001.

2- جولي بالانت (ت) خالد العامري، التحليل الإحصائي باستخدام برنامج SPSS، ط 2، دار الفاروق للنشر والتوزيع، مصر، 2009.

الجزء الخامس

الإحصاءات الوصفية المتعددة

- الفصل الخامس عشر: التوسع في جداول التقاطع: إضافة متغيرات التحكم
- الفصل السادس عشر: الارتباط الجزئي والانحدار المتعدد

الفصل الخامس عشر

التوسع في جداول التقاطع: إضافة متغيرات التحكم

مقدمة:

تناولنا في الفصل السادس تحليل العلاقة بين متغيرين، وقد تم افتراض أن أي علاقة تمتْ مشاهدتها في البيانات بين متغيرين هي ناشئة عن علاقة بسيطة ومباشرة. إن التطابق القوي في الجداول الثنائية، على أية حال، لا يعني بالضرورة أن هذه العلاقة البسيطة المباشرة موجودة واقعياً فالأمر يتعلق فقط، بالكيفية التي قمنا بها في تفسير هذه البيانات. ويمكننا القول، إن هناك علاقات أكثر تعقيداً تخفيها البيانات، لكننا لم نبذل جهداً كافياً لاكتشاف هذه العلاقات.

إن أسهل طريقة للتوسع في العلاقة، هي اكتشافها في جدول التوافق وذلك من خلال النظر إلى احتمالية تأثير متغير ثالث على العلاقة الثنائية الأصلية. واعتماداً على مخرجات هذا التوسع، فإنه يمكننا أن نعدل من نموذج العلاقة بين المتغيرين الأصليين آخذين في الاعتبار تأثير المتغير الثالث.

هناك ثلاث نتائج يمكن الوصول إليها عند إدخال المتغير الثالث في عملية التحليل:

1- النتيجة الأولى أنه ليس للمتغير الثالث أي تأثير على العلاقة الأصلية المباشرة.

2- قد يكون لإدخال المتغير الثالث تأثير في العلاقة الموجودة.

3- أو أن هناك علاقة مشروطة[1].

في هذا الفصل سوف نقوم باستقصاء هذه النتائج المحتملة، وذلك من خلال إعطاء أمثلة لكل نمط من أنماط هذه العلاقة.

1- العلاقة المباشرة:

دعنا ننظر إلى النقطة الأولى التي مفادها أنه لا يوجد أي تأثير لمتغير التحكم في العلاقة الثنائية الأصلية. ويبرهن لنا، هذا، على أن النموذج المباشر البسيط هو النموذج الملائم لوصف العلاقة.

على سبيل المثال، يمكننا إقامة علاقة بين الدخل ومشاهدة الإذاعة المرئية كما هي مرتبة في جدول التقاطع.. (انظر جدول 15 – 1).

جدول (15 ـ 1): مشاهدة الإذاعة المرئية حسب مستوى الدخل

مشاهدة الإذاعة المرئية	الدخل		المجموع
	منخفض	عال	
منخفض	115 % 57	95 % 32	210
عال	88 % 43	204 % 68	292
المجموع	203	299	502
Gamma = 0.47 = جاما			

إن العلاقة الأصلية في هذا النموذج بين هذين المتغيرين هي علاقة مباشرة كما في الشكل التالي:

مستوى الدخل ← مشاهدة الإذاعة المرئية

شكل (15 ـ 1) علاقة مباشرة

إن النموذج النظري الذي اعتمدنا عليه يوضح أن الدخل يؤثر مباشرة في كمية مشاهدة الإذاعة المرئية للشخص، وأن الإحصاءات تؤكد لنا أن هذه العلاقة الموجبة متوسطة إلى قوية، وعندما نجادل أن هناك علاقة مباشرة بين هذين المتغيرين فإننا بشكل فعّال نجادل في أن العلاقة سوف تكون نفسها بغض النظر عن أي متغير آخر يمكن أن يسبب تباين الحالات فيما بينها، في هذا المثال نعتقد أن الدخل يؤثر في مشاهدة الإذاعة المرئية بنفس الطريقة وبنفس الدرجة بغض النظر عن أي متغير آخر يمكنه أن يسبب التباين في الحالات مثل النوع، العمر... الخ.

إن هذه العلاقة الثنائية المباشرة على - أي حال - يمكن أن تظهر لنا أنها علاقة بسيطة. بالتأكيد توجد متغيرات أخرى قد تؤثر على كمية مشاهدة الإذاعة المرئية. فقد يشعر باحث آخر على سبيل المثال، بأن (المستوى التعليمي) يؤثر أيضاً في مشاهدة الأفراد للإذاعة المرئية. ولتقييم احتمالية تأثير هذا المتغير الجديد (المستوى التعليمي) أن له تأثير على العلاقة المشاهدة بين الدخل وكمية مشاهدة الإذاعة المرئية، يمكننا تقسيم العينة إلى مجموعتين فرعيتين:

المجموعة الأولى: تمثل أولئك الذين لم يكملوا التعليم الثانوي، والمجموعة الثانية: تمثل أولئك الذين أكملوا بعضاً من التعليم العالي. ومن الناحية الفنية فإننا قد أدخلنا متغير المستوى التعليمي كمتغير للتحكم. إن متغير التحكم هو ذلك المتغير الذي يحلل البيانات إلى مجموعات فرعية تستند على فئات متغير التحكم، إن تأثير متغير التحكم يؤدي بنا إلى توليد جداول متقاطعة منفصلة لكل المجموعات الفرعية والتي تم تحديدها من خلال متغير التحكم، في هذا المثال: دعنا نأخذ أولاً تلك (الحالات التي ليس لها تعليم عالٍ) ونصمم جدول تقاطع بين دخولهم ومشاهدتهم للإذاعة المرئية متغاضين عن تلك الحالات التي أكملت بعضاً من التعليم العالي، وبعد ذلك نأخذ فقط الحالات التي لديها بعضاً من

التعليم العالي ونصمم جدول تقاطع بين دخولهم ومشاهدة الإذاعة المرئية متغاضين عن أولئك الذين ليس لديهم تعليم عال.

إن ما قمنا به من توليد لهذه الجداول يطلق عليه الجداول الجزئية[2]. ويمكننا أن نولد من الجداول الجزئية العدد الذي نريده كفئات لمتغير التحكم (انظر جدول 15-2، 15-3) فإننا في هذين الجدولين قد أدخلنا المستوى التعليمي كمتغير للتحكم وتم قياس التعليم بفئتين فقط. وعليه يمكننا أن نولد جدولين جزئيين. (إذا كان لدينا ثلاث فئات لمتغير التحكم مثل: لا تعليم جامعي، بعض التعليم الجامعي، كثير من التعليم الجامعي في هذه الحالة يمكن توليد ثلاثة جداول جزئية).

جدول (15 ـ 2)

مشاهدة الإذاعة المرئية حسب مستوى الدخل

التحكم في مستوى التعليم (الذين ليس لديهم تعليم عالٍ)

مشاهدة الإذاعة المرئية	الدخل: منخفض	الدخل: عال	المجموع
منخفض	78 A %57	22 B %31	100
عال	58 C %43	48 D %69	106
المجموع	136	70	206
جاما = 0.49 = Gamma			

جدول (15 ـ 3)

مشاهدة الإذاعة المرئية

التحكم في مستوى التعليم (الذين أكملوا بعضاً من التعليم)

مشاهدة الإذاعة المرئية	الدخل		المجموع
	منخفض	عال	
منخفض	37 % 55	73 % 32	110
عـــــال	30 % 45	156 % 68	186
المجموع	67	229	296
جاما = 0.45 = Gamma			

وبهذه النتيجة يمكننا القول إن العلاقة الأصلية قد ولَّدت لنا تقريباً علاقة متساوية في كل جدول جزئي، وإن قيمة جاما لكل واحد من هذين الجدولين الجزئيين تقرب كثيراً من قيمة جاما في الجدول الأصلي قبل التحكم في مستوى التعليم. بمعنى آخر، بغض النظر عن المستوى التعليمي فإن العلاقة بين الدخل ومشاهدة الإذاعة المرئية ما زالت قائمة، إن العلاقة المباشرة التي شاهدناها في البداية لازالت باقية على ما هي عليه حتى بعد التحكم في المتغير الثالث. وبغض النظر عن كيف تتباين الحالات طبقاً لمستوى الدخل، فالعلاقة الثنائية المباشرة تبقى الشيء نفسه، وعليه فلن نغير من نموذجنا الأصلي الذي يوضح أن هناك علاقة مباشرة بين الدخل ومشاهدة الإذاعة المرئية[3].

إجراءات توليد جداول التقاطع مع التحكم في المتغيرات باستخدام SPSS:

1- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Analyze/ descriptive Statistics / Crosstabs (يعطيك مربع جداول التقاطع).

2- انقر على TV Watching (مشاهدة الإذاعة المرئية).

3- انقر على ◄التي تشير إلى القائمة المحددة المعنونة: Row (s) (الصف أو الصفوف).

نقوم بلصق (Pasts) TV Watching في القائمة المحددة للصف أو الصفوف Row (s)

4- انقر على Income.

5- انقر على ◄ التي تشير إلى القائمة المحددة المعنونة Column(s)

نقوم بلصق (Pasts) Income في القائمة المحددة للأعمدة Column(s).

6- انقر على مستوى التعليم Education Level.

7- انقر على ◄ التي تشير إلى القائمة المحددة أسفل Layer 1 of 1، نقوم بلصق (Pasts) Education Level في القائمة المحددة التي تحتوي المتغير المتحكم The Control Variable جدول التقاطع سوف يولد كل قيمة للمتغير في القائمة.

8- انقر على زر Statistics وقم باختيار Gamma هذه العملية ستولد قيمة جاما Gamma لكل جدول جزئي.

9- انقر على الخلايا Cells واختر نسب العمود Column Percentages.

10- انقر على Ok.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

Cross tabs TV Watching "Income" Education Level Crosstabulation

Education Level		Income		ToTAL
		Low	High	
No Post. Secondary	TV Watching Low Count	78	22	100
	% Within Income	57.4 %	34.4 %	48.5 %
	High Count	58	48	106
	% Within Income	42.6 %	88.6 %	51.5 %
	TOTAL Count	136	70	206
	% Within Income	100 %	100 %	100 %
Post. Secondary	TV Watching Low Count	37	73	110
	% Within Income	55.2 %	31.9 %	37.2 %
	High Count	30	156	186
	% Within Income	44.8 %	68.1 %	62.8 %
	TOTAL Count	67	229	296
	% Within Income	100 %	100 %	100 %

Symmetric Measures

Education Level			Value	Asymp. std. Error[a]	Approx.T[b]	Approx. Sig.
No Post. Secondary	ordinal by Ordinal	GAMMA	.492	.118	3.657	.000
	N of Valid cases		206			
Post. Secondary	ordinal by Ordinal	GAMMA	.450	.113	3.317	.001
	N of Valid cases		296			

a. Not assuming The null hypothesis.

b. Using The asymptotic standard Error assuming the null hypothesis.

المصدر: George Argyrous, Statistics for Social & Health Research, op.cit. , p. 465.

شكل (2-15) مخرجات SPSS لجداول التقاطع بإدخال متغير التحكم

تفسير مخرجات SPSS لجداول التقاطع مع التحكم في المتغيرات.

إذا نظرنا إلى هذه المخرجات نجد في الحقيقة جدولين متقاطعين جمعا في جدول واحد فالنصف الأول من الجدول، هو جدول التقاطع للدخل مع مشاهدة الإذاعة المرئية للحالات التي لم تتحصل على التعليم العالي. وتحت هذا الجدول مباشرة جدول التوافق لتلك الحالات التي أكملت بعضاً من التعليم العالي. إن نسبة الحالات التي تشاهد مستوىً معيناً من الإذاعة المرئية هي الشيء نفسه لكل فئات الدخل بغض النظر عن المستوى التعليمي. وقد عززت هذه النتيجة بقيم جاما Gamma كما ظهرت في الجدول المُعَنْوَن بـِ Symmetric Measures Table. وهذه القيم هي الشيء نفسه كالتي تم حسابها سابقاً. إن العلاقة بين الدخل ومشاهدة الإذاعة المرئية احتفظت بنفس القوة والاتجاه لكل جدول من الجداول الجزئية[4].

2- العلاقة الكاذبة أو الدخيلة:

نفترض أنه عندما أدخلنا المستوى التعليمي في التحليل تحصلنا بدلاً من الجدول (15-2 و15 -3) على جدول (15-4 و 15-5) من خلال هذين الجدولين (4 و 5) نلاحظ أن العلاقة بين الدخل ومشاهدة الإذاعة المرئية والتي شاهدناها في الجدول الأصلي قد اختفت فجأة. إنه بوضوح يمكننا أن نرى أنه لا يوجد ارتباط يمكن الحديث عنه بين الدخل ومشاهدة الإذاعة المرئية، وما إن أدخلنا مستوى التعليم كمتغير للتحكم حتى إن الارتباط الأصلي بين المتغيرين الدخل ومشاهدة الإذاعة المرئية قد اختفى بإدخال هذا المتغير وأن الطريقة الدقيقة للوصول لهذه النتيجة هي حساب جاما الجزئية من قيم جاما لكل جدول جزئي.

جدول (15 ـ 4) مشاهدة الإذاعة المرئية وفقا لمستوى الدخل: التحكم في مستوى التعليم (لأولئك الذين لم يتحصلوا على التعليم العالي)

مشاهدة الإذاعة المرئية	الدخل		المجموع
	منخفض	عال	
منخفض	102 % 75	50 % 71	152
عال	34 % 25	20 % 29	54
المجموع	136	70	206
جاما = 0.09 Gamma =			

جدول (15 ـ 5) مشاهدة الإذاعة المرئية وفقا لمستوى الدخل: التحكم في مستوى التعليم (الذين أكملوا بعضا من التعليم العالي)

مشاهدة الإذاعة المرئية	الدخل		المجموع
	منخفض	عال	
منخفض	13 % 19	45 % 20	58
عال	54 % 81	184 % 80	238
المجموع	67	229	296
جاما = −0.007 Gamma =			

جاما الجزئية Partial Gamma:

إن قيم جاما التي تم حسابها لكل جدول من الجدولين الجزئيين كانت ذات فائدة لكشف النقاب عن العلاقة الكاذبة أو الدخيلة. ويمكننا أن نرى أن هذه القيم المتعلقة

بجاما ضعيفة من حيث القوى باختلاف قيم جاما المشتركة بالجدول الثنائي الأصلي. في الجدول الأصلي، حيث إن الحالات لم تكن منفصلة وفقاً لمستوى التعليم، فقد وصلت قيمة جاما إلى 0.47. إلا أن قيم جاما لكل جدول من الجداول الجزئية تقرب من الصفر.

إن المنطق وراء حساب جاما الذي تَمَّتْ مناقشته في متن هذا الكتاب المتعلق بالأزواج المتوافقة والأعداد المتعلقة بالأزواج غير المتوافقة. فالأزواج المنسجمة أو المتوافقة كما بَيَّنَّا سابقاً هي مجموعة الحالات الزوجية التي تم ترتيبها بشكل متساوٍ مع كل متغير من هذين المتغيرين، وبالتالي شملت العلاقة الإيجابية بين هذين المتغيرين، بينما على الجانب الآخر، فإن الأزواج المتنافرة من الحالات والتي رتبت بشكل مختلف عن المتغيرين قد اشتملت على علاقة سالبة بين المتغيرين.

إذا ما أضفنا الأزواج المتقاربة أو المنسجمة عبر الجدولين الجزئيين وأضفنا الأزواج المتنافرة عبر نفس الجدولين فإنه باستطاعتنا حساب جاما الجزئية التي تقيس العلاقة المباشرة بين المتغيرين اللذين بدأنا بهما مع إدخال متغير ثالث كمتحكم، قد تم حساب جاما لجمع الأزواج المنسجمة والأزواج غير المنسجمة عبر الجداول الجزئية.

تجدر الإشارة إلى أننا لازلنا نستخدم كل الحالات لتحديد جاما الجزئية، ولكننا الآن نقوم بهذا العمل بعد أن تم فصل الحالات إلى جدولين جزئيين منفصلين. إن عملية حساب جاما الجزئية لهذه البيانات يمكن توضيحها في الجدول (15 – 6).

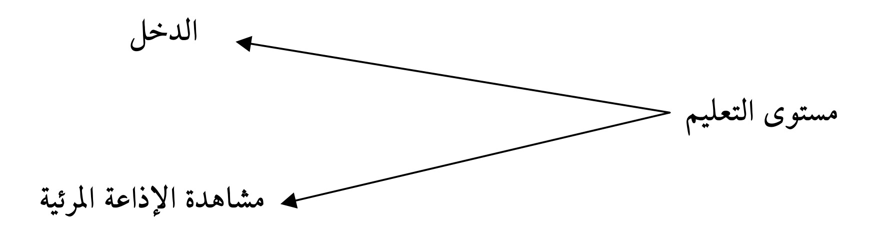
جدول (15 ـ 6) حساب جاما الجزئية

	الأزواج المنسجمة	الأزواج غير المنسجمة	جاما
الجدول الثنائي الأصلي	23.460 = 115 × 204	8360 = 95 × 88	0.47
الجدول الجزئي (1)	2040 = 102 × 20	1700 = 50 × 34	0.09
الجدول الجزئي (2)	2392 = 184 × 13	2430 = 45 × 54	-0.007
المجموع عبر الجداول الجزئية	4432 = 2392 + 2040	4130 = 2430+1700	0.04

إن قيمة جاما الجزئية لهذه البيانات هي فقط 0.04 مشيرة إلى وجود علاقة مباشرة ضئيلة بين الدخل ومشاهدة الإذاعة المرئية عندما تم إضافة متغير مستوى التعليم كمتغير متحكم(5).

العلاقة الكاذبة أو الدخيلة:

عندما كانت جاما الجزئية على درجة منخفضة إذا ما قورنت بقيمة جاما الأصلية التي تم حسابها في جدول التوافق المشترك يتوجب علينا أن نستنتج وجود إما علاقة كاذبة أو دخيلة بين المتغيرين الأصليين. وقبل تفسير كل نمط من أنماط هذه العلاقة، يحتم علينا الإشارة لأن نقرر أياً من النماذج يفسر النتائج المتعلقة بتوسيع الجداول هل هو تفسير نظري وليست مسألة إحصائية وعندما وجدنا أن العلاقة الأصلية قد اختفت بعد توسيع جداول التوافق، فالأمر يبقى لنا لنقرر كيف تتوافق هذه المتغيرات الثلاثة معاً استناداً إلى فهمنا إلى الواقع، فعلى سبيل المثال، يمكننا الاعتقاد أن النموذج الموضح أدناه (شكل 15 -3) هو أفضل تفسير للنتائج التي تم تحليلها الآن.



شكل رقم (15 ـ 3) علاقة كاذبة

إن هذا النموذج يوضح لنا العلاقة الكاذبة بين الدخل ومشاهدة الإذاعة المرئية بإدراك أن هذه العلاقة غير موجودة بين المتغيرين الأصليين وأن أن هذه العلاقة مجرد نتيجة إحصائية فقط استناداً على خصوصية علاقتهما بالمتغير المتحكم. فالمستوى التعليمي يؤثر بشكل منفصل على الدخل ومشاهدة الإذاعة المرئية، إلا أن المتغيرين الأخيرين ليسا مرتبطين ببعضهما البعض.

وقد نجد باحثاً آخر ينظر إلى هذه النتائج التي تحصلنا عليها من خلال توسيع جداول التقاطع بين الدخل ومشاهدة الإذاعة المرئية وبدلاً من ذلك يمكنه وصف العلاقة كما هي موضحة قي الشكل رقم (4).

الدخل ←——— المستوى التعليمي ←——— مشاهدة الإذاعة المرئية

شكل رقم (15 ـ 4) علاقة دخيلة

إن هذا الباحث يمكنه أن يجادل في أن أولئك الذين يكسبون دخلاً عالياً بإمكانهم الحصول على التعليم الجامعي، وهذا التعليم الجامعي يؤثر على كمية مشاهدة الإذاعة المرئية. ومهما كان اعتقادنا بأن هذه المناقشة جيدة أو غير ذلك، فهي مسألة ترجع في الأساس إلى كونها مسألة نظرية. أما إذا كان هذا التفسير أكثر ملاءَمةً لنتائج توسيع الجداول أكثر من نموذج العلاقة الكاذبة فهو أمر متروك لنقاش. لكن التحليل الإحصائي في حد ذاته لا يمكنه أن يقرر بين المسألة. فنتائج التحليل الإحصائي هي مجرد إشارة إلى أن واحداً من هذه النماذج يمكنه أن يفسر النتائج بشكل جيد.

العلاقة المشروطة:

نفترض أن باحثاً ما يرغب في معرفة إلى أي مدى يستجيب المرضى لبرنامج رياضي يهدف إلى تحسين نظام الأوعية الدموية (القلب) وقد قسم الباحث المرضى إلى مجموعتين: مجموعة ذات مستوى منخفض من الممارسة الرياضية والأخرى ذات مستوى عالٍ من الممارسة الرياضية ولاحظ ما إذا كان هناك أي تحسن في أنظمة الأوعية الدموية (انظر جدول رقم 7).

جدول (15 ـ 7) تحسن أنظمة الأوعية الدموية وفقا لمستوى الممارسة الرياضية

التحسن	المستوى الرياضي		المجموع
	منخفض	عال	
لا	38 % 73	11 % 34	49
نعم	14 % 27	21 % 66	35
المجموع	52	32	84

ومن خلال جدول (15 – 7) يمكننا أن نلاحظ أن خلايا النموذج لكل عمود يقترح أن هناك علاقة قوية وموجبة بين المتغيرين. إن البرنامج الرياضي يبدو أنه ذو فعالية. ولتعزيز هذه الفكرة فقد تم حساب جاما التي أنتجت قيمة تساوي 0.68.

إن الباحث يمكنه أن يتوقف عند هذه النتيجة، ويصل إلى قرار أن هناك علاقة مباشرة بين المتغير المستقل (مستوى التمارين الرياضية) والمتغير التابع (مستوى التحسن). وعلى أية حال، فإن الباحث يعتقد أن العلاقة الفعلية أكثر تعقيداً من هذا، إنه ربما توجد عوامل أخرى لم يضعها الباحث في هذا التحليل والتي يمكنها أن تقود إلى ما إذا كان قد تحسن نظام الأوعية الدموية (القلب) لدى المريض.

وبشكل خاص، يمكن للباحث أن يعتقد فيما إذا كان الشخص من المنتظمين في التدخين سوف يؤثر على فرصتهم للاستجابة لبرنامج الممارسة الرياضية. ومن هنا يتوجب على الباحث أن يعيد حساب جاما في جدول التوافق، وإدخاله متغير مستوى التدخين هذه المرة. والنتائج لهذا التحليل تظهر في الجدولين 8 و 9.

جدول (15 ـ 8): التحسن في الأوعية الدموية وفقا لمستوى الممارسة الرياضية (المدخنون فقط)

التحسن	مستوى الرياضة		المجموع
	منخفض	عال	
لا	28 % 74	7 % 70	35
نعم	10 % 26	3 % 30	13
المجموع	38	10	48
جاما = 0.09			

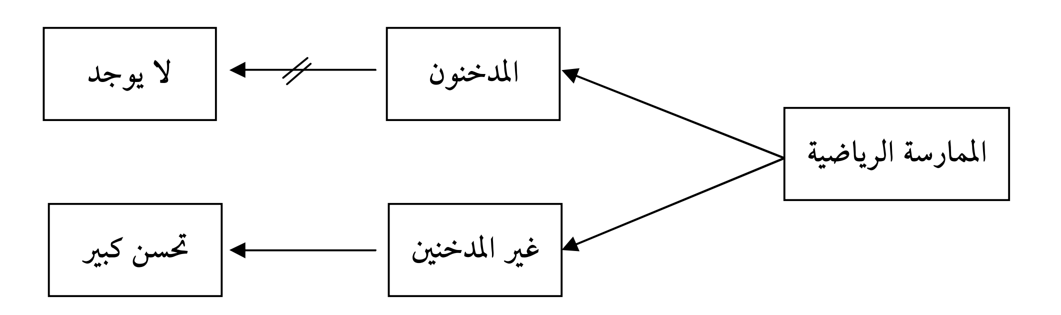
جدول (15 ـ 9) التحسن في نظام الأوعية الدموية وفقا لمستوى الممارسة الرياضية (غير المدخنين)

التحسن	مستوى الرياضة		المجموع
	منخفض	عال	
لا	10 % 71	4 % 18	14
نعم	4 % 29	18 % 82	22
المجموع	14	22	36
جاما = 0.84			

وعند مقارنة هذين الجدولين الجزئيين بالجدول الأصلي فإنه من الواضح أن العلاقة بين هذه المتغيرات تتباين اعتماداً على تاريخ الشخص المدخن. فالأفراد المنتظمين في التدخين لم يسجلوا أي تحسن في مستويات صحتهم كنتيجة لبرنامج الممارسة الرياضية. بينما سجل الأفراد غير المدخنين علاقة أقوى مما كانت عليه عند التعامل مع الجدول الأصلي. فالنتيجة قد ضعفت نتيجة لتضمين أولئك المدخنين مما أدى إلى عدم ثبات العلاقة وهذا يتأكد لنا

من خلال قيم جاما لكل جدول من هذه الجداول. فالجدول المتعلق بغير المدخنين فقط وصلت قيمة جاما إلى 0.84 مقارنة بـِ 0.68 للجدول الأصلي ككل. أما فيما يتعلق بالمدخنين، فإنه لا يوجد أي تحسن من برنامج الممارسة الرياضية، وبالتالي نجد أن النتيجة قد اختفت نتيجة لتأثير المتغير المتحكم. إن مقياس التطابق يعتبر مقيداً للغاية نظراً لأن مقياس التطابق يقيس التغيرات التي حدثت عندما أضيف متغير التحكم.

وكنتيجة لهذه الملاحظة فقد عمد الباحث إلى تغيير النموذج الذي يمكن أن يربط المتغيرات بعضها ببعض. وبدلاً من العلاقة البسيطة المباشرة ذات الاتجاه الواحد، فالباحث يمكنه أن يصف الارتباط في إطار العلاقة المشروطة كما هي مبينة في الشكل التالي:



شكل رقم (15 ـ 5) العلاقة المشروطة

إن العلاقة المشروطة يطلق عليها في بعض الأحيان (التفاعل)، فالتفاعل يكون موجوداً عندما تكون العلاقة بين متغيرين معتمدة على قيم خاصة بمتغير ثالث. وفي بعض الأحيان يمكننا أن نجد العلاقة عكسية اعتماداً على قيمة المتغير المتحكم؛ فقد تكون العلاقة لمجموعة فرعية علاقة موجبة، في حين يمكن أن تكون العلاقة سلبية في مجموعة أخرى فرعية.

مثال:

الآن يمكننا أن نستقصي العلاقة بين مقياس الذكاء (IQ) والدخل. فالذكاء مقياس وفقاً لمعيار اختبار IQ ويمكننا تقسيم المبحوثين إلى مجموعتين: مجموعة ذات مستوى ذكاء

منخفض والأخرى ذات مستوى ذكاء عالٍ. كذلك يمكننا تقسيم هؤلاء المبحوثين أيضاً إلى مجموعتين: مجموعة ذات دخل منخفض ومجموعة ذات دخل عالٍ وذلك طبقاً لمتوسط الدخل القومي لهذا المجتمع أو ذاك؛ إن النتائج المجمعة لـِ: 1000 شخص تم مسحهم يوضحها الجدول رقم (15 – 10)، ويوضح هذا الجدول أن هناك ارتباطاً متوسطاً بين الذكاء – كما تم قياسه باختبار IQ – والدخل، وقد تقود هذه النتيجة إلى تفسير إلى أن التباين في الذكاء يسبب التباين في مستوى الدخل. فقدرة الناس على الكسب تحدد من خلال ذكائهم الخاص. فهم إلى حد ما ومن أجل تحاشي مثل هذه النتيجة يمكننا أن نجادل بأن اختبار (IQ) كقياس للذكاء يكون اختباراً متحيزاً، إننا قد نشعر بشكل خاص أن درجات IQ هي انعكاس لخلفية الوضع الاجتماعي، وأن هذا المتغير هو الأساس في تحديد الدخل ولتقييم هذا الأمر يمكننا بناء جدولين جزئيين من خلال تقسيم الـ 1000 مبحوث إلى فئة اجتماعية عالية وفئة اجتماعية منخفضة، وإن هذه المجموعات الفرعية ولدت لنا النتائج التالية كما توضحها الجداول 15 – 10، 15 – 11، 15 – 12.

جدول (15 ـ 10) الدخل وعلاقته بالذكاء

الذكاء IQ	الدخل		المجموع
	منخفض	عال	
منخفض	165 %36	95 %18	260
عال	295 %64	445 %82	740
المجموع	460	540	1000
جاما = 0.45			

جدول رقم (15 ـ 11) الدخل وعلاقته بالذكاء
(الفئة الاجتماعية العُليا فقط)

الذكاء IQ	الدخل		المجموع
	منخفض	عال	
منخفض	20 %18	60 %14	80
عال	90 %82	380 %86	470
المجموع	110	440	550
جاما = 0.17			

جدول رقم (15 ـ 12) الدخل وعلاقته بالذكاء
(الفئة الاجتماعية الدنيا فقط)

الذكاء IQ	الدخل		المجموع
	منخفض	عال	
منخفض	145 %41	35 %35	180
عال	205 %59	65 %65	270
المجموع	350	100	450
جاما = 0.13			

ومن خلال هذه الجداول يتضح أن هناك علاقةً ثنائيةً قويةً قد اختفت بشكل كبير عندما تم إدخال الخلفية الاجتماعية كمتغير للتحكم. وكما أنه يمكن ملاحظة أنَّ هناك فرقاً بسيطاً في نمط التوزيعات النسبية عبر الجدولين الجزئيين. في حقيقة الأمر أنه عندما تم حساب جاما الجزئية على أساس الجداول الجزئية وجدنا أن قيمة جاما فقط 0.15 في هذه الحالة يمكننا القول بأنّ لدينا علاقة كاذبة أو علاقة دخيلة [6].

الخلاصة:

في هذا الفصل قد تمت مناقشة الطريقة التي من خلالها تم إدخال متغير ثالث يمكنه إحداث تغيير في العلاقة التي تم ملاحظتها سابقاً بين متغيرين. حقاً، قد يكون الأمر أكثر تعقيداً عندما يسمح بوجود تأثير متغيرات أكثر على العلاقة الثنائية الأصلية. آخذين في الاعتبار احتمالية تأثير متغيرات أخرى متعلقة بالتحليل المتعدد، والذي سوف نتناوله في الفصل اللاحق. ولمساعدة القارئ في فهم هذا الفصل يمكننا أن نرسم بعض النتائج من خلال توسعنا في جداول التوافق (انظر جدول 15-13) الذي يقدم لنا دليلاً لاتخاذ القرار.

جدول (15-13) النتائج المحتملة عند إدخال متغير التحكم

الجداول الجزئية عند مقارنتها بجداول التقاطع يتبيّن:	النموذج	التطبيقات لتحليل أبعد	احتمالية الخطوة التالية في التحليل الإحصائي	التطبيقات النظرية
العلاقات نفسها بين X و y	علاقة مباشرة	غُض النظر عن متغير التحكم	اختيار متغير تحكم آخر لزيادة اختبار العلاقة المباشرة	نموذج يشير إلى أن متغير X يسبب y في علاقة مباشرة تم دعمها
علاقة ضعيفة أو لا توجد علاقة بين X و y	علاقة كاذبة أو دخيلة	إدخال متغير التحكم أو إدخال متغير متحكم	• التركيز على العلاقة بين هذه المتغيرات الثلاثة • التركيز على العلاقة بين هذه المتغيرات الثلاثة	• النموذج يشير إلى أن X تسبب y لمتغير لم يتم دعمه • النموذج الذي يشير إلى أن X تسبب y جزئياً ولكن يجب تعديل النموذج لأخذ متغير التحكم في الاعتبار
علاقات مختلطة	التفاعل / علاقة مشروطة	إدخال متغير التحكم	تحليل المجموعات الفرعية استناداً على متغير التحكم بشكل منفصل	النموذج الذي يشير إلى أن X تسبب y ثم دعمه جزئياً، ولكن يتوجب علينا تعديل النموذج من أجل أخذ متغير التحكم في الحسبان

المصدر: Joseph F. Healey , Statistics: A Tool for Social Research , Wadsworth Cengage Learning , USA , 1993 , P. 428.

أسئلة للمراجعة:

1- ما هي النتيجة التي يمكنك استنتاجها حول العلاقة بين X و y استناداً على البيانات التالية:

كل الحالات:

Y	X: 1	X: 2	مجـ
1	177	146	323
2	51	346	397
مجـ	228	492	720

التحكم في C1:

Y	X: 1	X: 2	مجـ
1	153	52	205
2	44	123	167
مجـ	197	175	372

التحكم في C2:

Y	X: 1	X: 2	مجـ
1	24	94	118
2	7	223	230
مجـ	31	317	348

2- دراسة أوضحت العلاقة بين العمر، الاهتمام بالبيئة، والانتماء السياسي.. أنتجت لنا قيم جاما التالية:

- جاما (العمر مع الاهتمام بالبيئة): 0.57-
- جاما (العمر مع الاهتمام بالبيئة) (الليبراليين فقط): 0.22-
- جاما (العمر مع الاهتمام بالبيئة) (المحافظين فقط) 0.67-

ما هي النتيجة المستنتجة حول هذه العلاقة؟ إذا كانت هناك علاقة بين هذه المتغيرات.

الهوامش والمصادر:

أولا: الهوامش:

1- George Argyrous , Statistics for Social and Health Research With a Guide to SPSS , Sage Publications , London ,2001 , P.461

2- انظر: عبد الله عامر الهمالي، التقنيات الإحصائية ومناهج البحث، ط 1، منشورات جامعة قاريونس، بنغازي، 2008، ص ص 88-94.

3- George Argyrous , op.cit, PP. 462 - 463.

4- Ibid , PP. 464 - 465.

5- Ibid , PP. 466 - 467.

6- Ibid , PP. 470 - 471.

ثانيا: المصادر:

1- George Argyrous, Statistics for Social and Health Research With Guide to SPSS, Sage Publications, London, 2001.

2- Joseph F. Healey, Statistics for Social Research, Wadsworth Cengage Learning, USA, 1993.

3- ______________ , The Essentials of Statistics: A Tool for Social Research, Wadsworth Cengage Learning, USA, 2010.

الفصل السادس عشر
الارتباط الجزئي والانحدار المتعدد

مقدمة:

نحاول في هذا الفصل أن نتناول بعض التقنيات المستخدمة في تحليل العلاقات السببية التي تساعد الباحث في الوصول إلى تنبُّؤَات حول الظاهرة التي يسعى لدراستها. وتستند هذه التقنيات على ارتباط بيرسون (r) الذي تمتْ مناقشته في الفصل التاسع وتستخدم هذه التقنيات بشكل ملائم وعلى درجة عالية من الدقة، وتقاس متغيرات هذه التقنيات على مستوى المقياس ذي المسافات والنسبي.

إن أول هذه التقنيات التي سوف نناقشها في مستهل هذا الفصل هي تحليل الارتباط الجزئي.

أولاً: الارتباط الجزئي:

تعريفه: الارتباط الجزئي هو عبارة عن تقنيات إحصائية تسمح للباحث أن يفحص العلاقة الثنائية بين متغيرين، بينما يتحكم في المتغير الثالث، والذي يمكن الإشارة إليه في هذا الشأن بـِ (A) أو متغير التحكم.

حساب الارتباط الجزئي:

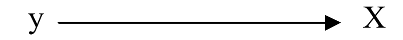
لحساب الارتباط الجزئي - بادئ ذي بدء - ينبغي على الباحث إجراء ارتباط بيرسون (r)، وبعد ذلك يمكنه الدخول في إجراءات حساب الارتباط الجزئي. وتجدر الإشارة هنا إلى الاختلاف بين الارتباط الثنائي (r) والارتباط الجزئي، حيث يفترض الباحث أن هناك تأثيراً للمتغير الثالث على العلاقة الثنائية بين المتغيرين تحت الاستقصاء. فعلى سبيل المثال، إذا كان الزوج المتعلم، والزوج غير المتعلم لديهما استجابات مختلفة فيما يتعلق بعدد الأطفال في الأسرة. فإن الارتباط الجزئي سيختلف في القوة، وربما في الاتجاه إذا قورنا بمعامل الارتباط الثنائي.

وقبل الدخول مباشرة في حساب معامل الارتباط الجزئي، دعنا نأخذ بعين الاعتبار العلاقات بين معامل الارتباط الجزئي ومعامل الارتباط الثنائي، وماذا تعني هذه الارتباطات حيث توجد ثلاثة أنماط محتملة من هذه العلاقات.

أنماط العلاقات[1]:

1- العلاقات المباشرة:

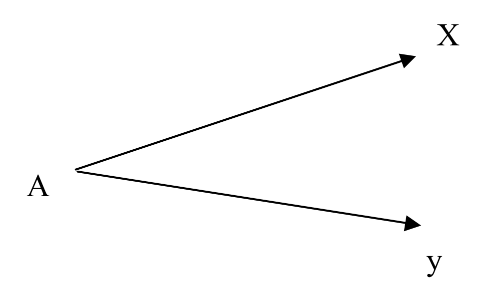
في الحقيقة إنه عندما يُدخل الباحث متغير التحكم في تحليل بياناته، فإنه قد يتحصل على أحد الاحتمالات، وهي أن معامل الارتباط الجزئي في الأساس قيمته مساوية للمعاملات الثنائية. فعلى سبيل المثال، إذا تم التحكم في تعليم الزوج، فإننا نجد أن قيمة معامل الارتباط الجزئي تصل إلى 0.49+ إذا قورنت بارتباط بيرسون (r) 0.50+. وهذا يعني إن إدخال المتغير الثالث لا تأثير له على العلاقة الأصلية (عدد الأطفال وعدد الساعات المتعلقة برعاية هؤلاء الأطفال). بمعنى آخر، أنه بغض النظر عن مستوى التعليم، فإن الأزواج يستجيبون بطريقة واحدة لمسألة رعاية الأطفال. إن مثل هذه النتيجة تكون متوافقة مع النتيجة التي تشير إلى العلاقة المباشرة أو السببية بين X و y.



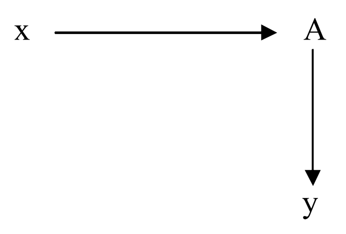
وأن المتغير الثالث لا علاقة له بالاستقصاء. وفي هذه الحالة ينبغي على الباحث أن يغض النظر عن المتغير الثالث (A) في التحليلات الإحصائية اللاحقة.

2- العلاقات الكاذبة والدخيلة:

إن الاحتمال الآخر الذي يمكن أن يحدث هو عندما تكون قيمة معامل الارتباط الجزئي ضعيفة إذا ما قورنت بمعامل الارتباط الثنائي، وربما يصل هذا الارتباط إلى صفر. وقد تكون النتيجة متساوية مع علاقات مختلفة بين المتغيرين. أولهما العلاقة الكاذبة Spurious Relationship. فالتحكم في متغير (A) هو سبب في وجود كلا المتغيرين، المتغير المستقل (X) والمتغير التابع (y)، كما هو موضح في الشكل التالي:



وتعني هذه النتيجة أنه لا توجد علاقة حقيقية بين X و y، ويظهر هذان المتغيريان أنهما مرتبطان بعضهما ببعض فقط بوجود متغير التحكم A. وعندما يضع الباحث في اعتباره متغير التحكم (A)، فإن العلاقة المرئية بين X و y ستختفي. وثانيهما العلاقة الدخيلة بين X و y هذا النمط (يكون الارتباط الجزئي أكثر ضعفاً من معامل الارتباط الثنائي) متسقاً Consistent مع العلاقة الدخيلة بين المتغيرين كما هو موضح في الشكل التالي:



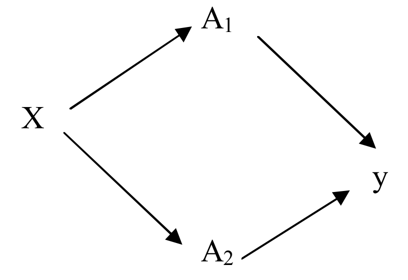
في مثل هذا الموقف، نجد أن X و y متغيران لا يرتبطان بعضهما ببعض مباشرة

ولكنهما – سببياً – مرتبطان بوجود التحكم (A). مرة أخرى، إنه عند التحكم في متغير (A) فإن العلاقة المرئية بين X و y ستختفي.

إن السؤال الذي يمكن طرحه في هذا الخصوص هو: كيف يمكن للباحث أن يميز بين العلاقة الكاذبة والعلاقة الدخيلة؟ للإجابة عن هذا السؤال، يمكننا القول، إن التمييز بينهما لا يمكن القيام به من خلال الأسس الإحصائية: فالعلاقة الكاذبة والعلاقة الدخيلة هما شيء واحد من الناحية الإحصائية: فالباحث يمكنه أن يميز بين هاتين العلاقتين فقط من حيث ترتيب المتغيرات، من حيث الوقت (بمعنى أي من المتغيرين قد جاء أولاً) أو يمكنه التمييز بين هاتين العلاقتين على أسس نظرية *Theoretical grounds* وليس على أسس إحصائية.

التفاعل **Interaction**:

وأخيراً، فإن الاحتمال الثالث للعلاقة بين المتغيرين يمكن الإشارة إليه هنا، بالرغم من صعوبة اكتشافه من خلال تحليل الارتباط الجزئي، ويطلق على هذه العلاقة: التفاعل. وتحدث علاقة التفاعل بين X و y وتتغير بشكل ملحوظ تحت قيم متنوعة لـِ A. فعلى سبيل المثال، قد تكون هناك علاقة موجبة بين X و y لأحد فئات A، وعلاقة سالبة للفئة الأخرى كما هو مبين في الشكل التالي:



حساب وتفسير معامل الارتباط الجزئي:

المصطلحات والمعادلات: تتطلب معادلة الارتباط الجزئي بعض المصطلحات الجديدة. وسنتعامل هنا مع أكثر من واحد من العلاقات الثنائية، وبالتالي، نحتاج أن نميز بين هذه

العلاقات من خلال رموز معينة. وعليه، فإن الرمز ryx سوف يشير لمعامل الارتباط بين المتغير y والمتغير x، و ryA يشير إلى معامل الارتباط بين متغير y ومتغير A، و rxA يشير إلى معامل الارتباط بين X و A.

معامل الارتباط الجزئي:

إن أول علاقة جزئية يمكن الإشارة إليها بالرموز الجبرية التالية:

ryx.A

يمثل المتغير A إلى اليمين من النقطة، المتغير المتحكم وعليه، ryx.A تشير إلى معامل الارتباط الجزئي التي تقيس العلاقة بين المتغير x والمتغير y مع التحكم في متغير A. والمعادلة المتعلقة بالارتباط الجزئي:

$$ryx.A = \frac{ryx - (ryA)(rxA)}{\sqrt{1 - r^2}\, yA\sqrt{1 - x}\,A}$$

ملاحظة: ينبغي على الباحث أن يجري أولاً معامل الارتباط بين كل زوجين من المتغيرات (المتغيرات X و y و X و A، و y و X) قبل استخدام هذه المعادلة.

حساب معامل الارتباط الجزئي:

لتوضيح حساب معامل الارتباط الجزئي يمكننا استخدام المثال التالي:

مثال: البيانات التالية تمثل التدخين، الإصابة بالزكام، والإجهاد:

X التدخين	Y الإصابة بالزكام	A الإجهاد
70	2	4
105	6	3
35	1	2
105	4	3
0	2	2
70	7	3
35	3	1
140	6	5
0	4	5
140	4	3
$\overline{X} = 70$	$\overline{y} = 3.9$	$\overline{A} = 3.10$
$S^2 X = 2450.0$	$S^2 y = 3.49$	$S^2 A = 1.49$
$SX = 49.50$	$Sy = 1.87$	$SA = 1.22$

المصدر: George Diekhoff , Statistics for Social and Behavioral Sciences , Univariate , Bivariate , Multivariate , wm. c. Brown Publishers ,USA , 1992 , PP. 257 - 258

rxy =.53

rAy =.40

rAx =.29

$$ryx.A = \frac{ryx - (ryA)(rxA)}{\sqrt{1 - r^2 yA}\sqrt{1 - r^2 xA}}$$

$$= \frac{(.53) - (.40)(.29)}{\sqrt{1 - (.40)^2}\sqrt{1 - (.29)^2}}$$

$$= \frac{0.414}{\sqrt{0.84}\sqrt{0.92}}$$

$$= \frac{0.414}{(0.92)(0.96)}$$

$$= \frac{0.414}{0.883}$$

$$ryx.A = 0.47$$

التفسير: إن هذه القيمة التي تقيس العلاقة بين التدخين والإصابة بالزكام مع التحكم في متغير الإجهاد هي أقل من معامل الارتباط (r) ryx.53.

إن مربع الارتباط الجزئي هو 22. = $rxy.A = .47^2$. وتشير هذه الدرجة إلى أن 22% من التباين للتعرض للإصابة بالزكام يمكن التنبؤ به من خلال عدد مرات التدخين عندما يتم التحكم في متغير الإجهاد المهني(2).

اختبار الدلالة للارتباط الجزئي:

إن الارتباط الجزئي بين التدخين (X) والإصابة بالزكام (y) والتحكم في الإجهاد المهني (A) تم حسابها rxy.A =.47.

إن السؤال الذي يمكن طرحه هنا هو: ما إذا كانت هذه القيمة للارتباط الجزئي كافية. بمعنى هل هذه القيمة من الارتباط الجزئي دالة إحصائياً؟

دعنا نرى ذلك من خلال اتّباع الخطوات التالية:

الخطوة الأولى: صياغة الفرض الصفري والفرض البديل:

Ho: إن القيمة المشاهدة لِـ rxy.A مرجعها إلى خطأ المعاينة. إن الارتباط الجزئي بهذا الحجم نسبياً يكون مرجحاً أنه يوجد في عينة سحبت من مجتمع يكون فيه قيمة rxy.A تساوي صفراً. $rxy.A = 0$

Hi: إن القيمة المشاهدة لـ rxy.A كبيرة جداً لأن تكون مرجحة لخطأ المعاينة. إن هذه القيمة من غير المحتمل، أن العينة المسحوبة من مجتمع تكون فيه rxy.A تساوي صفراً

وتعطي درجة عالية للارتباط الجزئي. وعليه، فإن احتمال العينة قد جاء من المجتمع الذي تكون فيه 0 < rxy.A، وأن هناك احتماليةً كبيرةً بأنه في حالة إعادة الدراسة فإن درجة الارتباط الجزئي تكون أكبر من صفر (0).

الخطوة الثانية: إحصاء الاختبار The Test Statistic:

إن الارتباط الجزئي rxy.A يساوي 47. يمثل إحصاء الاختبار للعلاقة بين X و y بعد التحكم في متغير A.

الخطوة الثالثة: تحديد الاحتمالية المرتبطة بإحصاء الاختبار.

بمقارنة القيمة المحسوبة لِ rxy.A المساوية لـِ 47. لتوزيع المعاينة rxy.A نستطيع عندئذ rxy.A تحديد الاحتمالية القريبة بأن العينة التي تعطي هذه القيمة من الارتباط الجزئي قد جاءت من مجتمع فيه تكون قيمة rxy.A تساوي صفراً (0). ولدرجة حرية df=N-J=10-3=7، حيث (N) تشير إلى عدد الحالات و(J) تشير إلى عدد المتغيرات الخاضعة للتحليل، وباعتماد اختبار أحادي الجانب، فإن القيمة الحرجة لـِ (r) كما هو موضح في جدول توزيع قيمة (r) تحت مستوى دلالة 05. تكون القيمة مساوية لـِ rxy.A تساوي 47. تقع تحت القيمة الحرجة، فإننا بالتالي نصل إلى نتيجة مفادها أن الاحتمالية أكبر من 0.05. إن العينة المسحوبة من مجتمع تكون فيه rxy.A تساوي صفراً (0) ستعطينا ارتباطاً جزئياً يصل إلى 47. = rxy.A، وعليه فإن العينة المدروسة إلى حد كبير يرجح أنها تمثل هذا المجتمع. بمعنى آخر، إن العلاقة بين التدخين والإصابة بالزكام ليست دالة إحصائياً عند التحكم في الإجهاد المهني إحصائياً[3].

الارتباط الجزئي باستخدام SPSS:

1- إدخال البيانات Data Entry:

إدخال قيمة X في العمود الأول، وقيم y في العمود الثاني و A في العمود الثالث في محرر بيانات Spss.

2- تحليل البيانات Data Analysis:

1- انقر على Anayze ⟸ واختر Correlate وانقر على Partial.

2- في الوقت نفسه يلقى الضوء على قائمة العمود لـِ X و y على الجانب الأيسر من العمود ويتم تحريكها إلى مربع المتغيرات.

3- أَلْقِ الضوء على قائمة العمود لـِ A ويتم التحرك إلى مربع التحكم.

4- انقر على OK.

3- مخرجات Spss:

تمدنا هذه المخرجات بمصفوفة الارتباط Correlation Matrix، الارتباطات المحتملة بما فيها الارتباط بين X و X، والارتباط بين y و y. نحن نرغب في الارتباط الجزئي بين X و y التي يكون في أعلى زاوية من يمين المصفوفة. وتحت الارتباط يظهر مستوى الدلالة Significance Level. وعندما يكون مستوى الدلالة أقل من 05. فإن ذلك يشير إلى أن معامل الارتباط معامل دالة.

ثانياً: تحليل الانحدار المتعدد:

مراجعة الانحدار الثنائي Bivariate Regression:

في المثال التالي يمكننا أن نستقصي العوامل التي تؤثر في قيمة مبيعات ما. لقد قامت إحدى الوكالات للخدمات العقارية بجمع بيانات حول متغيرين لعدد 12 منزلاً. وقد جاءت النتائج وفقاً للجدول التالي:

جدول (16 ـ 1) حجم المنزل وأسعار البيع

سعر البيع ($،000)	حجم المنزل	سعر البيع ($،000)	حجم المنزل
260	20	287	24
240	18	252	20
245	20	270	23
210	13	275	25
230	18		
242	14		
295	28		
235	16		

المصدر: George Argyrous, Statistics for Social & Health Research, with a Guide to SPSS , SAGE Publications, London ,2001 , p. 475

إن الغرض من هذا التحليل هو تفسير التباين في سعر بيع هذه المنازل التي تم التعامل معها كمتغير تابع. وتعتقد وكالة الخدمات العقارية أن العامل الأساسي المفسر للتباين في أسعار البيع يرتبط بالتباين في مساحات هذه المنازل. وبالتالي يمكن أن نطلق عليه نموذج العوامل المحددة لسعر المنزل. وبما أن هذا النموذج هو نموذجُ نظريٌّ يصف العلاقة التي يمكن أو لا يمكن أن تصمد أمام التدقيق الأمبيريقي.

دعنا نقارن، على سبيل المثال، منزلين من عينة هذه المنازل، كالمنزل الذي تم بيعه بقيمة 252,000 ألف دولار، بالمنزل الآخر الذي تم بيعه بقيمة 230,000 ألف دولار. من خلال هذه المقارنة، يمكننا القول، بأن قيمة البيع مرتبطة بحجم المنزل. ومن هنا نجد أن هذين المنزلين يبدو أنهما متناغمان مع نموذج هذه الوكالة للخدمات العقارية.

إن السؤال الذي يمكن طرحه هنا هو: هل هذه العلاقة تبقى صحيحة في كل المنازل الأثنى عشر؟

وبتطبيق تحليل الانحدار البسيط لهذه البيانات الواردة في الجدول أعلاه مستخدمين طريقة أقل المربعات المألوفة Method of Ordinary Least Square، تتولد لدينا النتائج التالية:

$$y = 157 + 4.88 X$$

$$r = 0.92$$

$$r^2 = 0.85$$

وعلى ضوء هذه النتائج يمكننا التوصل إلى:

- هناك علاقة موجبة بين حجم المنزل وسعر البيع.
- أن أي زيادة في حجم المنزل تقود إلى الزيادة في سعره بحوالي 4880 دولاراً.
- إن هذه العلاقة، علاقة قوية وبدرجة عالية من الثقة لإجراء عملية التنبؤ.
- إن التباين في حجم المنزل ليس بالضرورة أن يكون على نحو كامل للتنبؤ بسعر المنزل، فمعامل التحديد، معامل قيمته عالية (0.85)، ولكنها ليست مساوية لواحد صحيح. ومع ذلك يمكن القول، بأن هناك عوامل أخرى تؤثر في سعر عينة هذه المنازل.

إن ثمن البيع الفعلي لأي منزل من هذه المنازل يمكننا التعبير عنه بالمعادلة التالية:

ثمن البيع = a + b (حجم المنزل) + e

وتشير هذه المعادلة إلى أن ثمن مبيعات هذه المنازل تتباين بشكل أساسي بسبب الفروق في أحجامها، وكذلك بسبب عوامل عشوائية تمثلت من خلال مصطلح الخطأ (e). ويشير مصطلح الخطأ (e) إلى الفرق بين ما تم التنبؤ به فيما يتعلق بثمن المنزل من خلال حجمه المحدد، والثمن الذي بَيْع به فعلاً.

وتجدر الإشارة إلى أن هناك عدة عوامل تؤثر في مبيعات هذه المنازل، إلا أن النموذج المعتمد لدى وكالة الخدمات العقارية ترى أنه بالرغم من العوامل الأخرى التي تؤثر في مبيعات هذه المنازل، إلا أنه من بين هذه العوامل لا يوجد متغير واحد (حجم المنزل) الذي يلعب دوراً أساسياً في تحديد ثمن البيع بطريقة نظامية وثابتة. وهذا ما حدا بنا إلى التركيز على متغير حجم المنزل وإعطائه وضعاً بيّناً في المعادلة. إلا أنه مع ذلك يمكننا القول، أيضاً إننا لا نريد أن نتجاهل العوامل الأخرى المؤثرة في ثمن بيع هذه المنازل.

إن مصطلح الخطأ (e) يخدم كل هذه العوامل الأخرى، فالعوامل التي تؤثر في ثمن بيع المساكن في كيفما اتفق وبطريقة غير نظامية. فقد يباع أحد المنازل بقيمة عالية، لأن وكالة الخدمات العقارية كانت، بشكل خاص، تطرح هذه المنازل للبيع بطريقة مغامرة ووضع ثمن معين لهذا المنزل، في حين يباع منزل آخر، لأن المشتري قد جذبه بشكل خاص نظام الألوان؛ في حين، يمكن أن يباع منزل آخر بثمن بخس لأن البائع Vendor يريد بيع المنزل بشكل سريع نظراً للدين محدد السداد، ونتيجة لهذه العوامل وعوامل أخرى، تنبثق عشوائياً من بيع إلى بيع آخر الذي لا ينبغي علينا أن نتعامل معها كمتغيرات مستقلة ومنفصلة. ولكن بالسماح لمصطلح الخطأ أن ينتزع تأثير هذه العوامل مجتمعة. ففي بعض الأحيان، هذه العوامل العشوائية تُسَبِّبُ زيادةً عاليةً في عملية البيع أكثر مما نتنبأ به استناداً على المعرفة المتعلقة بحجم المنزل، وفي أحيان أخرى تتسببَّ هذه العوامل في انخفاض ثمن البيع بحيث يكون أقل من القيمة المتوقعة. وبمعرفة حجم المنزل سوف يُسْمَح لنا بالتنبؤ بقيمة البيع التي سوف تكون قريبة من الثمن المستهدف. إلا أننا نسلّم بأن منزلاً معيناً فإن تأثير هذه العوامل العشوائية سوف تعني أن ثمن البيع الحقيقي ليس بالضرورة أن يكون مساوياً للثمن الذي تمَّ التّنُّبؤ به[4].

مقدمة للانحدار المتعدد:

تجدر الإشارة إلى أنه يمكن أن ننظر إلى النموذج الثنائي Bivariate Model بشكل مبسط أكثر مما ينبغي. فقد يتولد لدينا شعورُ بأن هناك عواملَ أخرى غيرَ عامل (حجم المنزل) قد لا تكون عواملَ عشوائيةً، ولكنها عواملُ تعملُ بطريقةٍ نظاميةٍ مُسَبِّبَةٍ التباينَ في قيمة مبيعات المنازل وتباين هذه العوامل بشكل مستقل عن حجم هذه المنازل. بمعنى آخر، إذا قارنا منزلين لديهما نفس الحجم، فالفرق في ثمن البيع لهذين المنزلين قد لا يكون راجعاً فقط للعوامل العشوائية التي ناقشناها للتو. فإذا نظرنا إلى الجدول رقم (1). فإننا نجد أن هناك ثلاثةَ منازلَ في عينة الدراسة - على سبيل المثال - لديها نفس الحجم. فقد تم بيع أحد هذه المنازل بِـ 260,000 ألف دولاراً، والآخر بِـ 245,000 ألف دولاراً، والثالث بِـ 252,000 ألف دولاراً. فالسؤال الذي يمكن طرحه هنا هو لماذا هذه الفروق في أسعار البيع؟ بالرغم من أن هذه المنازل متساوية من حيث الحجم؟ إذا وضعنا ثقتنا

التامة في النموذج الثنائي والقول بأنَّ العوامل العشوائية مسئولة عن تفسير هذه الفروق. أو يمكننا القول، بأنَّ نموذجاً آخَرَ يمكن أن يسمح لفعالية متغيرات أخرى بشكل نظامي للتأثير على سعر البيع خَرَجْنَا بتفسيرات أفضل.

ويمكننا، على سبيل المثال، الاعتقاد بأن عمر المنزل أيضاً (وبشكل جزئي) يفسّر هذا التباين في أسعار هذه المنازل. بمعنى، أن عمر المنزل ليس متغيراً أحياناً يمكن أن يؤثر على ثمن بيع المنزل، ولكن بدلاً من ذلك، يعتبر عامل شائع وأنه على نحو منتظم يؤثر على ثمن بيع المنازل.

إن النموذج الجديد يمكن أن يظل قائماً بشكل معقول لتوقع أنه كلما زاد عمر المنزل قل سعره. بمعنى، أننا نتوقع أن هناك علاقة سلبية بين سعر المنزل وعمره. وإذا ما اعترانا الشك فإن ذلك كذلك، فإن الأمر يتطلب منا أن نوسع من تحليل الانحدار ليشمل هذا التحليل العملية المتعلقة بهذا المتغير الآخر، وهذه العملية تشبه إلى حد كبير نفس الطريقة التي اتبعناها في الفصل الخامس عشر المتعلق بالتوسع في التحليل الثنائي لجداول التقاطع والأخذ في الاعتبار التأثير المتوقع للمتغير الثالث. وعند التعامل مع بيانات المقياس ذي المسافات والنسبي كما هو الحال الآن يكون الهدف من وراء تحليل الانحدار المتعدد *Multiple Regression*. فالانحدار المتعدد هو ذلك التحليل الذي يستقصى العلاقة بين مجموعة المتغيرات المستقلة (متغيرين أو أكثر) ومتغير تابع.

واستناداً على هذا النموذج المتعدد، توضح البيانات الواردة في الجدول (16 – 2) عمر الاثنى عشر منزلاً.

جدول (16 ـ 2) ثمن البيع، حجم المنزل، عمر المنزل (12 منزلاً)

ثمن البيع ($0.000)	حجم المنزل	عمر المنزل (بالسنوات)
260	20	5
240	15	12
245	20	9
210	13	15
230	18	9
242	14	7
295	28	1
235	16	12
287	24	2
252	20	5
270	23	5
275	25	5

المصدر: George Argyrous, Statistics for Social & Health Research, op.cit , p. 478

بشكل عام، يمكننا التعبير عن العلاقة بين أي متغير تابع، وأي عدد من المتغيرات المستقلة (K) بالطريقة التالية:

$$y = a + b_1 x_1 + b_2 X_2 + \dots b_k x_k + e$$

ومن خلال المثال الذي بين أيدينا، فإنه يمكننا أن نوضح نموذج العلاقة بالتعبيرات التالية:

(ثمن البيع) a + b_1 (حجم المنزل) + b_2 (العمر) + e

بمعنى آخر، إننا نعتقد بأن ثمن بيع المنزل ينطلق في اتجاه واحد أو آخر وفقاً لعمر وحجم المنزل. وإننا بذلك نتوقع أن المنزل القديم ذا الحجم الصغير نسبياً سيئُول سعره نحو الاتجاه الأسفل عبر العملية المستقلة لكلا المتغيرين العمر والحجم. وبشكل عكسي، فإننا نتوقع أن منزلاً حديثاً وكبير الحجم نسبياً سيتجه سعر بيعه نحو الأعلى. وفي شواهد أخرى، فإن حجم وعمر المنزل من الممكن أن يتجها في اتجاهات معاكسة.

وبالرجوع إلى المعادلة السابقة. فإننا لازلنا نسمح للعوامل العشوائية أن يكون لها تأثير. وبالتالي فإن عمر وسعر المنزل لا يحددان تماماً ثمن البيع في كل حالة. ولكن إذا ما كان هذا النموذج المتعدد Multivariate Model هو الأفضل الذي يمكن الركون إليه في تفسير قيمة بيع المنازل، إذا ما قورن بالنموذج الثنائي Bivariate Model، فإن كمية التباين قد تركت لكي تفسر من خلال مصطلح الخطأ (e)، ستكون أصغر في النموذج المتعدد إذا ما قورنت بالنموذج الثنائي الذي بدأنا به في مستهل هذا الفصل. إلا أنه على الجانب الآخر، فإن إدخال العمر في المعادلة لم يكن قادراً على التقليل من نسبة التباين في ثمن البيع الذي يعزى إلى الخطأ (e). إذاً من خلال معرفتنا بعمر المنزل، ليس بالضرورة، أن يُحسِّن من قدرتنا بالتنبؤ بسعر بيعه. ومن هنا نكون قد أضعنا الوقت والجهد في جمع معلومات حول متغير لا فائدة ترجى منه.

إن مهمة الانحدار المتعدد هي أن يوزع التباين في سعر المنازل لكل واحد من هذه العوامل المتنافسة في التأثير على المتغير التابع، هل يهيمن أحد هذه العوامل على تحديد سعر البيع، كقولنا إن عمر أو حجم المنزل يكون بشكل واضح أكثر أهمية. أم أن هذين العاملين يكونا لديهما نفس التأثير؟

والسؤال الآخر الذي يمكن طرحه في هذا السياق، ما الدور الباقي للعوامل العشوائية؟ إن التحليل المتعدد، وفقاً لحساب معامل الانحدار أو الارتباط الجزئي لكل متغير يمدنا بمقاييس دقيقة للتأثير الخاص لهذه المتغيرات المستقلة على المتغير التابع. إنه من الممكن استخدام تقنيات عددية لحساب معامل الانحدار بين كل واحد من هذه المتغيرات المستقلة والمتغير التابع. ومع ذلك، فإن هذه التقنيات ستكون مُرْهِقَةً بشكل كبير ومُسْتَهْلِكَةً للوقت، خاصة إذا كان الباحث يتعامل مع مجموعة كبيرة من البيانات. وتجدر الإشارة هنا إلى أنه، ومع تقدم تقنيات الحاسوب، أصبح من السهولة بمكان على الباحث حساب الانحدار المتعدد من خلال البرامج المعدة لذلك مثل SPSS الذي يقوم بإنهاء العمليات الحسابية بشكل سريع ودقيق. وتبقى مهمة الباحث الأساسية هي تفسير هذه النتائج[5].

إجراءات الانحدار المتعدد باستخدام SPSS:

إن إجراء حساب الانحدار المتعدد يشبه إجراء حساب الانحدار الثنائي الذي تناولناه في الفصل التاسع من هذا الكتاب، ما عدا أننا في تحليل الانحدار المتعدد نقوم بلصق Paste أكثر من متغير إلى المتغيرات المستقلة Independent (s) إلى القائمة المستهدفة Target List.

الإجراء:

1- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Analyze ⇐ Regression ⇐ Linear

(يعطيك صندوق الحوار لِـ Linear Regression Dialog box)

2- انقر على Selling Price (سعر البيع) في قائمة المتغير Variable List.

3- أنقر على ◄ الذي يشير إلى المتغير التابع Dependent: القائمة المحددة للمتغير. هذه العملية ستقود إلى لصق (Pastes) Selling Price كمتغير تابع.

4- انقر على House Size (حجم المنزل) في قائمة المتغير، وبينما تضغط على مفتاح Shift، انقر على العمر بالسنوات Age in years.

5- انقر على ◄ الذي يشير إلى المتغير المستقل (المتغيرات) Independent (s) في القائمة المحددة للمتغيرات. هذه العملية ستقود إلى عملية لصق كلا المتغيرين House Size وعمر المنزل بالسنوات Age in years كمتغيرين مستقلين.

6- انقر على OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

Variables Entered / Removed[b]

Model	Variables Entered	Variables Removed	Method
I	Age in years House size (Squares[a])		ENTER

a. All Requested Variables Entered

b. Dependent Variable: Selling Price ($0.000)

Model Summary

Model	R	R Square	Adjusted R Square	Std Error of the Estimate
I	.959[a]	.919	.901	.807

a. Predictor: (Constant) , age in years , House size (Squares)

Anova[b]

Model	Sum of Squares	Df	Mean Square	F	Sig.
I Regression Residual Total	.7596248 .158548 .9176796	2 9 11	24.37931 .90660	.29851	.000

a. Predictors (Constant) , age in years , House size (in Squares)

b. Dependent Variable: Selling Price ($0.000)

Coeffi cients[a]

Model	Unstandarized coefficients		Standariz ed coeffi cients	T	Sig.
	B	Std Error	Beta		
I (Constant)	.290224	.22226		.5538	.000
House size (Squares)	.5782	.973	.487	.6502	.026
Age in years	-2.974	.0761	-.503	-2.764	.022

a. Dependent Variable: Selling Price ($0.000)

المصدر: George Argyrous, Statistics for Social & Health Research, op.cit, p. 481.

شكل رقم (16 ـ 1) مخرجات SPSS للانحدار الخطي

إذا نظرنا إلى هذه المخرجات، فإننا نجد كثيراً من المعلومات قد تولدت من خلال برنامج SPSS، وعليه، ينبغي على الباحث أن يركز على أهم الأجزاء في هذه المخرجات. فالجدول المُعَنْوَنُ بـِ Variables Entered / Removed، متغيرات أدخلت / أبعدت، يبين لنا وصفاً لفظياً بسيطّاً للنماذج التي تمكننا من التكهن.

إن برنامج SPSS، بإمكانه أن يقوم بحساب عدة انحدارات متعددة بشكل تلقائي، وذلك من خلال استخدام عدة مجموعات من المتغيرات المستقلة ليحدد أيًّاً من هذه المجموعات أو التوليفات تعطينا أفضل تفسير للتباين في المتغير التابع. هنا في هذا المثال، تنبأنا بنموذج واحد أطلق عليه نموذج (1)، الذي استخدم فيه متغير العمر بالسنوات وحجم المنزل بالمربعات كمتنبئ للمتغير التابع قيمة البيع (000.$).

أما الجدول الثاني، والذي أطلق عليه خلاصة النموذج Model Summary فإنه يبين لنا قيمة معامل الارتباط التي تشير إلى قوة العلاقة بين مجموعة المتغيرات المتوالفة في النموذج Model، والمتغير التابع. وتشير قيمة R إلى 959.، علاقة قوية. أن قيمة R هي الارتباط المتعدد مرادفاً لمعامل الارتباط الثنائي، (r). وأن R^2 (معامل الانحدار المتعدد للتحديد) تصل قيمتها إلى 919. وعندما يتم استخدام النموذج الثنائي Bivariate Model لتفسير سعر البيع (عندما تجرى عملية الارتباط فقط مع حجم المـنزل House Size) فإن قيمة معامل التحديد قد وصلت إلى 0.85. وعندما نستخدم كلا المتغيرين: حجم المنزل وعمره للتنبؤ بسعر البيع، فإن معامل التحديد ترتفع لتصل إلى 919.. وتشير هذه القيمة إلى قدرتنا على التفسير (أو التنبؤ) بقيمة بيع المنزل التي زادت عندما توفرت لدينا معلومات حول عمر وحجم المنزل. إن جزءاً من التباين في قيمة البيع التي أرجعناها سابقاً إلى عوامل عشوائية، هي في حقيقة الأمر تعزو للتأثير النظامي لعمر المنزل.

أما الجدول المُعَنْوَن بـِ ANOVA أنوفا فهو جدولُ يحتوي على إحصائيات استدلالية inferential Statistics، تسمح لنا بإجراء استدلالات من خلال العينة على المنازل الأثنى عشرة. إلى هذه النقطة فإن التركيز يدور حول الإحصاءات الوصفية للعينة، وعليه، ينبغي على الباحث أن يتجاوز هذا الجزء من هذه المخرجات. والنظر إلى الجدول المُعَنْوَنِ بـِ Coefficients الذي يوفر عناصر معادلة الانحدار Elements of the regression equation

التي تم التنبؤ بها والتي يمكن كتابتها بالشكل التالي:

(حجم المنزل بالمربعات) 2.578 + 224.29 = قيمة البيع (\$.000)

(عمر المنزل بالسنوات) 2.974

وعند قراءة معادلة الانحدار، فإنه من الأهمية بمكان، أن نضع في أذهاننا وحدات القياس التي على ضوئها تم قياس المتغيرات. ففي المثال الذي بين أيدينا أن أي زيادة في كل مربع واحد في حجم المنزل تؤدي إلى الزيادة في سعر المنزل بقيمة تصل إلى 2578 دولاراً. وبالتحرر من هذه العلاقة، فقد وجدنا أيضاً أن زيادة كل سنة في عمر المنزل تقلل من قيمة البيع (لاحظ العلاقة السالبة) على نحو 2974 دولاراً.

دعنا الآن نتعامل مع شيء عملي. نفترض أننا نود أن نشتري منزلاً حجمه 15 متراً مربعاً وعمره خمس سنوات. ما هي توقعاتنا بالقيمة المطلوبة لبيعه؟ نضع المعلومات في معادلة الانحدار:

(\$.000) قيمة البيع = 224.29 + 2.578 (15) − 2.974 (5)

$= 224.29 + 38.67 - 14.87$

$= 248.09$

بطبيعة الحال، فإننا لا نتوقع أن يكون 248,090دولاراً الثمن الحقيقي الذي يتحقق عندما يتم بيع المنزل فعلياً، حيث إن العوامل العشوائية لازالت تلعب دوراً كبيراً في هذا الشأن. ولكن، وبوجود درجة عالية لمعامل التحديد، فإن هذه العوامل العشوائية ينبغي ألا تكون علة الثمن الفعلي للانحراف كثيراً عن القيمة المتوقعة.

إنه من الصعوبة بمكان استخدام معامل الانحدار لنقيّم الأهمية النسبية لكل متغير مستقل في تحديد قيمة المتغير التابع طالما أن كل متغير مستقل قد تم قياسه بوحدات مختلفة (أحدهما تم قياسه بعدد السنوات، والآخر بالمربعات). وإذا قمنا بقياس حجم المنزل بوحدة أخرى كالقدم المربع، حينئذٍ ستكون قيمة معامل الانحدار لهذا المتغير مختلفة وذلك

لاختلاف وحدة القياس التي تم اعتمادها في قياس هذا المتغير. بمعنى آخر، لا يمكننا القول بأن ذلك بسبب معامل حجم المنزل التي وصلت إلى 2.578. في حين أن قيمة المعامل المتعلقة بالعمر، قد وصلت إلى 2.974-. من هنا نجد أن متغير العمر متغيرُ قوي وله تأثير فاعل على قيمة البيع. وعمود المعامل المعيارية Standardized Coefficients توضح ذلك. وبدون الغوص في تفاصيل كيف تم حساب هذه المعامل المعيارية (يطلق عليها أيضاً أوزان بيتا beta - weights ، فإننا ببساطة نلاحظ أن هذه الأوزان (المعامل) قد أبعدت تأثير وحدات القياس. إنه باستطاعتنا أن نرى أن العمر (508.-) لديه قوة تأثير طفيفة على قيمة البيع، أكثر من التأثير الذي يحدثه حجم المنزل (487.).

الجدول التالي يلخص الدور الذي تلعبه المقاييس المختلفة التي يولدها برنامج SPss:

جدول رقم (16 ـ 3) تفسير مخرجات SPSS

معامل الانحدار ← (Regression Coefficients)	تسمح لنا بإجراء التنبؤ على المتغير المعياري (التابع) استناداً على قيم المتغيرات المستقلة في إطار الوحدات الأصلية للقياس.
المعامل المعيارية ← (Standardized Coefficients)	تسمح لنا أن نميز بين الأهمية النسبية لكل متغير في تحديد قيمة المتغير التابع.
R ←	تشير إلى قوة العلاقة بين مجموعة المتغيرات المستقلة (كتلة واحدة) والمتغير التابع (المعياري).
R^2 ←	تشير إلى كمية التباين في المتغير الذي تم تفسيره خلال مجموعة المتغيرات المستقلة في النموذج (1)، بذلك إشارة إلى ما إذا كان النموذج (1) نموذجاً جيداً للتنبؤ بالمتغير المعياري.

المصدر: George Argyrous, Statistics for Social & Health Research, op.cit, p. 482.

اختبار الدلالة للنموذج المتعدد:

إن المعلومات الحرجة لاختبار الاستدلال يوضحه الجدول المعنون بـِ ANOVA. ويقوم برنامج SPSS بإجراء اختبار F على النموذج بكامله لاختبار الفرضية المتعلقة بمعامل الارتباط Correlation Coefficients لكل المتغيرات التي يحويها النموذج Model. وهي صفر (0). في هذا المثال الذي بين أيدينا، فإن إحصاء اختبار F لهذا النموذج يكون مستوى الدلالة فيه 0.000 وهذا بين لنا أنه على الأقل أن أحد الارتباطات بين كل واحد من المتغيرات المستقلة والمتغير المعياري ليست مساوية لصفر في المجتمع. وتؤكد هذه النتيجة ما تم الحصول عليه في جدول معامل (Coefficients). حيث إن قيمة t الإحصائية لكل متغير مستقل أنها ذات دلالة على مستوى 0.05.

وعليه، فإنه باستخدامنا لاختبار F (F-test) نرى ما إذا كان على الأقل بعض المتغيرات المستقلة في النموذج هي دالة، وأن إحصاءات t (t-Stataitics) لكل متغير على حدة يشير إلى أي من هذه المتغيرات دالة إحصائياً[6].

الانحدار التدرجي Stepwise Regression:

قد يلاحظ مكتب الخدمات العقارية أن هناك متغيرات أخرى ذات أهمية كتلك المتغيرات التي تعاملنا معها، تلعب هي الأخرى دوراً أساسياً في تحديد قيمة بيع تلك المنازل في هذه المنطقة. وبالرغم من القوة التفسيرية للنموذج ذي المتغيرين المستقلين، فإن مكتب الخدمات العقارية يمكنه أن يجادل في أن قدرتنا في التنبؤ بأسعار هذه المنازل ستكون أفضل عند إدخال متغير مساحة الأرض المقام عليها المنزل، كمتغير آخر مستقل. وعليه، فإن مكتب الخدمات العقارية سيقوم بجمع معلومات إضافية حول مساحة الأرض المقامة عليها هذه المنازل الأثنا عشر. والجدول التالي يوضح هذه البيانات الإضافية.

جدول (16 ـ 4) سعر البيع، حجم المنزل، عمر المنزل والمساحة المقام عليها المنزل (12 منزلاً)

سعر البيع ($0.000)	حجم المنزل (مربعات)	عمر المنزل (بالسنوات)	مساحة الأرض المقام عليها المنزل (بالمتر المربع)
260	20	5	420
240	15	12	640
245	20	9	600
210	13	15	590
230	18	9	700
242	14	7	720
295	28	1	624
235	16	12	590
287	24	2	710
252	20	5	630
270	23	5	700
275	25	5	710

المصدر: George Argyrous, op.cit , p. 483.

الآن، لدينا ثلاثة نماذج لتفسير سعر بيع هذه المنازل. إن أول هذه النماذج: النموذج الثنائي الذي أخذ في اعتباره حجم المنزل، وعلاقته بسعر البيع. وثاني هذه النماذج النموذج المتعدد الذي أخذ في اعتباره حجم وعمر المنزل وعلاقتهما بسعر البيع. في حين يركز النموذج المتعدد الثالث على متغيرات حجم المساحة كمتغير تفسيري:

نموذج (1) سعر البيع = $a + b_1$ (حجم المنزل) + e

نموذج (2) سعر البيع = $a + b_1$ (حجم المنزل) + b_2 (عمر المنزل) + e

نموذج (3) سعر البيع = $a + b_1$ (حجم المنزل) + b_2 (عمر المنزل) + b_3 (مساحة الأرض) + e

إننا لِلتّوِّ قَدْ ناقشنا الطريقة التي من خلالها يمكن الحكم ما إذا كان متغير ما يضيف إلى القوة التفسيرية للنموذج من خلال النظر إلى تأثيره الضمني على قيمة مربع R. فإذا زادت قيمة R تربيع بشكل دالٍ عندما أضيف متغير إلى النموذج. حينئذٍ فالمعلومات الإضافية التي وفرها لنا هذا المتغيِّر المضاف زادت من قدرة النموذج في تفسير التباين في سعر البيع.

تجدر الإشارة إلى أن هناك طريقةً واحدةً لنقرر من خلالها بين عدة نماذج، وهي الشروع في إجراء انحدارات خطية منفصلة استناداً على مجموعة مؤتلفة من المتغيرات المستقلة التي نود أن يشملها النموذج. وبعد ذلك، يمكننا أن نقارن قيم R تربيع لنرى إلى أي مدى، تكون أيٌّ من قدرتنا قادرةً على تفسير التباين في سعر البيع الذي يصل إلى مداه الأعلى عن طريق كل واحد من هذه المجموعة المؤتلفة من المتغيرات المستقلة. على سبيل المثال، إذا قمنا بإجراء الانحدار المتعدد مضافاً إلى هذه الأجزاء مساحة الأرض، فإن قيمة R تربيع ستصل إلى 0.922 وهي قيمة مساوية للقيمة التي تحصلنا عليها في النموذج (2). الذي يحتوي على متغيرين فقط هما عمر المنزل وحجمه. بمعنى آخر، إن مساحة الأرض المقام عليها المنزل لا تؤدي إلى زيادة قدرتنا في تفسير سعر البيع، فالوقت والجهد في قياس هذا المتغير، هما في واقع الأمر عمل ضائع[7].

إن المشكلة الكامنة في هذه الطريقة، أنها طريقة مملة في القيام بإجراء انحدارات منفصلة لكل واحد من هذه النماذج الممكن بناؤها. إضافة إلى ذلك، صعوبة الحكم على مقدار الزيادة في R تربيع لتبرير ما يتضمنه أي متغير في النموذج. إلا أنه من حسن الطالع أن برنامج SPSS قادر على أن يمدنا بطريقة يطلق عليها الانحدار التدرجي حيث تسمح لنا هذه الطريقة بتحديد أي من المجموعة المتآلفة من المتغيرات المستقلة تقدم لنا أفضل تفسير للمتغير المعياري. وتتم هذه العملية بإضافة وحذف متغيرات من العملية الحسابية طبقاً إلى ما إذا كان كل واحد من هذه المتغيرات يحدث دلالة إحصائية تكون سبباً في تغير قيمة R تربيع[8].

إجراءات الانحدار المتعدد التدرجي باستخدام SPSS:

1- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Analyze ⟸ Regression ⟸ Linear

(تقود هذه العملية إلى مربع الحوار (Linear Regression)

2- انقر على Selling Price (سعر البيع) في قائمة المتغير.

3- انقر على ◄ التي تشير إلى Dependent في القائمة المحددة للمتغير (هذه العملية تقود إلى لصق Selling Price (Pasts) (كمتغير تابع).

4- انقر على House Size (حجم المنزل) في قائمة المتغير، مع الإبقاء ضاغطاً على مفتاح Shift والنقر على Age in years (العمر بالسنوات) وبعدها انقر على Land Size (مساحة الأرض).. (هذه العملية تلقي الضوء على House Size , Age in years and Land Size).

5- انقر على الـ ◄ التي تشير إلى المتغيرات المستقلة (s) Independent في القائمة المحددة للمتغيرات. (هذه العملية تقود إلى لصق Pasts، House Size , Age in years , Land Size كمتغيرات مستقلة).

6- انقر على ◄ القريبة من Enter (تظهر قائمة Adrop - down menu).

7- في القائمة drop-down menu انقر على Stepwise (تقود هذه العملية إلى اختيار الانحدار التدرجي Stepwise regression كطريقة لتضمن وتبعد المتغيرات من الانحدار.

8- انقر على OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء :

Regression

Variables Entered / Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Age in years		Stepwise (criteria) Probability-of-f-to Enter=.050 Probability-of-f-to remove=1.00
2	House size (Squares)		Stepwise (criteria) Probability-of-f-to Enter=.050 Probability-of-f-to remove=.100

a. Dependent Variable: Selling Price ($0.000)

M0del Summary

Model	R	R Square	Adjusted R Square	Std Error of the Estimate
1	.925a	.856	.842	.999
2	.959b	.919	.901	.807

a. Predictors: (Constant) , age in years.

b. Predictors: (Constant) , age in years , House size (Squares).

ANOVA[b]

Model	Sum of Square	Df	Mean Square	F	Sig
1 Regression	.9995820	1	.9995820	.84859	.000 a
Residua	5.91897	10	.59297		
Total	6.917879	11			
2 Regression	.7596248	2	24.37831	.29851	.0006 b
Residua	.158548	9	.90880		
Total	.9176796	11			

a. Predictors (Constant) , age in years.

b. Predictors (Constant) , age in years , House size (Squares)

c. Dependent Variable: Selling Price ($0.000)

Coeffi cients[a]

Model	Unstandarized coefficients		Standarized coefficients	T	Sig.
	B	Std Error	Beta		
1 (Constant)	.702292	.8325		.19350	.000
Age in years	-5.419	702	-.925	-7.723	.000
2 (Constant)	.290224	.22226		.5638	.000
Age in years	-2.974	.0761	-.509	-2.764	.022
House size (Squares)	.5782	397	.487	.6502	.026

a. Dependent Variable: Selling Price ($0.000)

Excluded Variables[a]

Model	Beta IN	t	Sig.	Partial correlation	Collinearit Statistics
					Tolorance
1 House size (Squares)	.487a	.6502	.026	.662	.265
Land size in meters squares	.015	.115	.911	.038	.973
2 Land size in meters squares	.015b	.147	.886	.052	.973

a. Predictors in the model (Constant) , age in years.

b. Predictors in the model (Constant) , age in years , House size (Squares)

c. Dependent Variable: Selling Price ($0.000)

المصدر: George Argyrous, Statistics for Social & Health Research, op.cit , p. 486.

شكل (16 ـ 2): مخرجات SPSS للانحدار الخطي

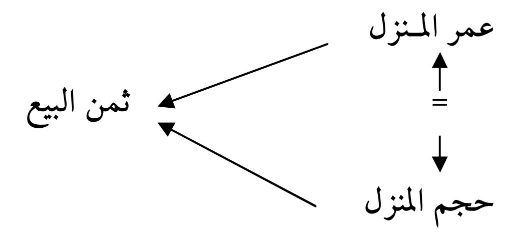
ولتفسير هذه المخرجات:

الجدول الأول المُعَنْوَن بِـ متغيرات / أدخلت وأبعدت Variables Entered / Removed في هذا الجدول تم توليد نموذجين من ثلاثة متغيرات تم افتراضها: الأول متضمناً متغير العمر بالسنوات (النموذج الثنائي الأساسي) والذي أطلق عليه في برنامج SPSS ، نموذج (1) (Model 1)، والنموذج الآخر يشتمل على متغير العمر وحجم المنزل (بالمربعات)، والذي أطلق عليه في برنامج SPSS نموذج (2) (Model 2)، أما باقي المخرجات فهي أساساً نفس المخرجات السابقة التي تم توليدها بشكل منفصل من قِبل كل واحد من هذه النماذج، وهنا لدينا نموذجان تم تقديمهما بنفس التحليل. أما الجزء الجديد من المخرجات فيوضحه الجدول الأخير المُعَنْوَنُ بِـ "Excluded Variables"، المتغيرات المستثناه. ويوضح هذا الجدول، وعلى أساس اختبار F وعلى أساس التغيرات في R^2. إن مساحة الأرض المقام عليها المنزل ليست متغيراً مهماً يمكن أن يضمن في أي من هذه النماذج Models. ويعني هذا كله، أن تعقيد النموذج بإضافة هذا المتغير الجديد لن يُؤَمن لنا أي نوع من الدقة فيما يتعلق بتقدير المتغير التابع، وبالتالي ينبغي الاستغناء عنه[9].

الافتراضات التي ينبغي مراعاتها عند استخدام الانحدار المتعدد:

على الرغم من قوة الانحدار المتعدد كأداة قوية لتقييم تأثير مجموعة من المتغيرات المستقلة على المتغير المعياري، إلا أن هناك مجموعةً من الافتراضات حول هذا التحليل تقلل من تطبيقاته. إن كل الافتراضات التي تمت مناقشتها عند الحديث عن الانحدار الثنائي فهي قائمة في تطبيقها على حالة الانحدار المتعدد. إضافة إلى هذه القائمة فإننا بحاجة إلى إضافة افتراض آخر مهم وهو أن التحليل المتعدد يفترض أن كل متغير من المتغيرات المستقلة مستقل عن الآخر. أي تسامت متعدد Mutticollinearity (أي التعدد الذي يقع على نفس الخط).

في المثال الذي بين أيدينا يمكن توضيحه بالشكل التالي:



إن كلاً من عمر وحجم المنزل يؤثران في سعر بيعه، ولكنهما لا يؤثران في بعضهما البعض. وقد يبدو لنا أن هذا الأمر معقول لهذه المتغيرات تحديداً: فإذا تم بشكل مفاجئ توسيع المنزل، فإن هذا الأمر لا يقود بشكل مفاجئ لأنْ يجعل من هذا المنزل منزلاً عتيقاً أو جديداً! وبشكل مشابه إذا ما زاد عمر المنزل، لا يعني بالضرورة أن يزيد حجمه أو يصغر. إن هذه الفرضية تحتوي ضمناً الانحدار المتعدد، مما يجعله إلى حد ما أكثر تعقيداً من التقنيات المتعددة الأخرى التي تناولناها في الفصل السابق، حيث تم استخدام التحليل المتعدد، لتحديد أي من هذه النماذج يقدم لنا أفضل تفسير للعلاقة بين المتغيرات الثلاثة أو أكثر منها[10]. ومن خلال تحليل الانحدار، فإننا نفترض نموذجاً محدداً كما تم في مخرجات SPSS شكل رقم (1)[11].

حدود استخدام الانحدار المتعدد والارتباط:

يعتبر الارتباط الجزئي، والارتباط المتعدد، تقنيات إحصائية قوية لتحليل العلاقات المتبادلة بين ثلاثة أو أكثر من المتغيرات. وتسمح هذه التقنيات للباحث بالقدرة على التنبؤ بدرجات متغير واحد من اثنين أو أكثر من متغيرات أخرى.

ولما كانت هذه التقنيات على درجة عالية من القوة فهي بالتالي تتطلب جهداً كبيراً، ودرجة عالية من البيانات، وقياساً على المستويين ذي المسافات والنسبي، وهذه متطلبات، عادة يصعب على الباحث إنجازها. إضافة إلى ذلك، فإن هذه التقنيات تستند على افتراضات: أن العلاقة المتبادلة بين المتغيرات تتبع شكلاً محدداً.

أولاً، تفترض هذه التقنيات أن كل متغير مستقل لديه صبغة العلاقة الخطية مع المتغير التابع. ولكي يتعرف الباحث على هذه العلاقة الخطية عليه القيام بمراجعة لشكل الانتشار.

ثانياً: تفترض هذه التقنيات أن تأثير المتغيرات المستقلة هي تأثيرات إضافية additive. وهذا يعني، أنه ينبغي علينا أن نفترض أفضل تنبؤ للمتغير التابع (y) يمكن الحصول عليه، وذلك من خلال إضافة درجات المتغير المستقل بعضِها إلى بعضٍ، على سبيل المثال (الفقر عندما يوجد مع الانعزالية) يولد درجة عالية من معدلات الجريمة.

ثالثاً: أن تقنيات الانحدار المتعدد، ومعامل الارتباط تفترض أن المتغيرات المستقلة متغيرات غير مرتبطة. هذه الحالة تعني أن معامل الارتباط بين كل أزواج المتغيرات المستقلة ينبغي أن تكون صفراً. في الممارسة العملية، ومع ذلك نعمل كما لو أن هذا الافتراض قد تم تحقيقه، إذا كانت الارتباطات المتبادلة بين المتغيرات المستقلة على درجة منخفضة.

وأخيراً، يجب الإشارة إلى أننا في هذا الفصل قد غطينا فقط التطبيقات البسيطة للارتباط الجزئي، والانحدار المتعدد، من حيث المنطق والتفسير، ولزيادة التفصيل لهذه التقنيات عندما يكون لدينا متغيرات مستقلة كثيرة واضحة المعالم نسبياً، ومع ذلك، فإن حسابات مثل هذه المواقف تعتبر معقدة جداً. وعندما يواجه الباحث مثل هذه المواقف لمتغيرات تزيد عن ثلاثة حينئذٍ يتطلب منه الاستعانة ببرنامج SPSS أو SAS للتعامل مع هذا الكم الهائل من المتغيرات[12].

أسئلة للمراجعة:

1- من البيانات التالية:

المدينة	الاحتجاج (y)	معدل البطالة (x)	% الدعاية السلبية
A	55	5	60
B	60	8	63
C	65	9	55
D	68	9	53
E	70	10	44
المتوسط	63.6	8.2	$\bar{x}$ = 55.8
الانحراف	5.5	1.7	S = 5.3
			% الدعاية السلبية
الاحتجاج	0.95 معدل البطالة		−0.87
معدل البطالة	–		−0.70

المطلوب:

a) حساب الارتباط الجزئي بين y و x، والتحكم في A ما هو التأثير الذي يحدثه متغير التحكم على العلاقة الثنائية. هل العلاقة بين الاحتجاج ومعدلات البطالة علاقة مباشرة.

b) احسب معامل الارتباط الجزئي بين (y) الاحتجاج، و (x) الدعاية السلبية متحكماً في معدل البطالة (A). ما هو التأثير الذي يحدثه متغير التحكم على العلاقة الثنائية. هل العلاقة بين الاحتجاج والدعاية السلبية علاقة مباشرة؟

c) اكتب ملخصاً قصيراً يبين نتائج هذه العلاقات بين المتغيرات الثلاثة.

2- من البيانات التالية: احسب معامل الارتباط الجزئي لعدد 12 حالة، مستخدماً اختبار أحادي الجانب:

rxy =.79

rAx =.81

rAx =.72

3- قام باحث بدراسة العوامل التي من الممكن أن تقود إلى التقليل من عدد أيام العمل والتي فقدت نتيجة للمرض، في مصنع معين. عشرة أشخاص تمت دراستهم. والبيانات التالية حول عدد الساعات التي يمارسون فيها التمارين الرياضية في الأسبوع، وكذلك أيام العمل التي غابوا فيها نتيجة لتعرضهم للمرض، يوضحها الجدول التالي:

عدد ساعات التمرين	عدد الأيام المفقودة	العمر بالسنوات
3	12	36
8	10	35
1	10	54
0	15	42
0	18	41
4	7	25
7	7	32
2	14	39
5	9	43
0	16	29
9	8	32
3	10	50

المطلوب:

- بين المتغير المعياري.
- ماذا تتوقع أن تكون الإشارة أمام المتغيرات المستقلة.
- ما هي معادلة الانحدار؟
- احسب معامل الانحدار المتعدد.
- هل تضمين العمر في النموذج يضيف شيئاً لقدرتنا على التنبؤ بعدد الساعات التي فقدت بسبب المرض.

4- بين كيف تم تطبيق تحليل الانحدار التدرجي في مجال البحث الاجتماعي، وما هي أسس هذا التطبيق؟

الهوامش والمصادر:

أولا: الهوامش:

1- Joseph F. Healey , The Essentials of Statistics: A Tool for Social Research , Wadsworth Cengage Learning , USA , 2010. pp. 362 - 364.

2- George DiEkhoff, Statictics for social and Behavioral Sciences: Univariat, Bivariate, Multivarite, wm. c. Brown Publishers , USA , 1992 , P. 261.

3- Ibid , P. 263.

4- George Argyrous , Statistics for Social and Health Research With Guide to SPSS , Sage Publications ,London , 2001 , P. 477.

5- Ibid , P. 479.

6- Ibid , P. 483.

7- Ibid , P. 484.

8- Ibid , P. 484.

9- Ibid , P. 485.

10- Ibid , P. 487.

11- Ibid , P. 487.

12- Joseph F. Healey , op. Cit. PP. 375 , 377 , 379.

ثانيا: المصادر:

1- George Argyrous , Statistics for Social and Health Research With Guide to SPSS , Sage Publications , London , 2001.

2- George Diekhoff, Statistics for Social and Behavioral Sciences, Univariate, Bivariate, Multivariate, wm. c. Brown Publishers, USA, 1992.

3- Joseph F. Healey, The Essentials of Statistitics: A Tool for Social Research, Wadsworth Cengate Learning, USA. 2010.

4- جولي بالانت (ت) خالد العامري، التحليل الإحصائي باستخدام برامج SPSS، ط 2، دار الفاروق للنشر والتوزيع، القاهرة، 2009.

5- عبد الله عامر الهمالي، التحديث الاجتماعي: معالمه ونماذج من تطبيقاته، منشورات جامعة قاريونس، 2008 م.

الجزء السادس

الإحصاءات اللابارامترية

- الفصل السابع عشر : اختبار مربع كاي ($X2$)لفروق الدلالة لعينة واحدة لتوزيع تكراري
- الفصل الثامن عشر : اختبار مربع كاي ($X2$) للاستقلال
- الفصل التاسع عشر : اختبار توزيع ذي الحدين لعينة واحدة
- الفصل العشرون : الأساليب الإحصائية للبيانات الترتيبية : اختبارات : مان وتني ، ولكوكسن ، كروسكال وليز وفريدمان

الفصل السابع عشر

اختبار مربع كاي (X^2) لفروق الدلالة لعينة واحدة لتوزيع تكراري

مقدمة:

إن كل الاختبارات التي تم تناولها في هذا الكتاب صممت في الأساس لاختبار الفروض حول معلمات محددة للمجتمع الإحصائي. فعلى سبيل المثال، لقد تم استخدام اختبارات t لتقييم الفروض حول u، ولاحقاً حول $u_1 - u_2$. إضافة إلى ذلك، أن هذه الاختبارات، على نحو نموذجي تقوم بافتراضات حول شكل توزيع المجتمع وحول معلمات أخرى للمجتمع. وبالرجوع بالذاكرة إلى تحليل الأنوفا ANONA، فإن توزيع المجتمع قد تم افتراض أنه موزع توزيعاً طبيعياً (أي اعتدالية التوزيع)، ومتجانس التباين، هي عملية مطلوبة، هي الأخرى، في هذا التحليل، ولما كانت هذه الاختبارات تهتم بالمعلمات وما تتطلبه هذه المعلمات من افتراضات، فقد أطلق عليها الاختبارات البارامترية Parametric Tests.

ومن الخصائص العامة الأخرى للاختبارات البارامترية أنها تتطلب وجود درجة عددية لكل حالة في العينة المدروسة. وتضاف الدرجات لبعضها البعض، ثم تُرَبَّع، ويُؤْخَذُ لها المتوسط. وبطريقة أخرى، تعالج باستخدام العمليات الحسابية الأساسية. أما

فيما يتعلق بموازين القياس، فإن الاختبارات البارامترية تتطلب بيانات مقاسة على المستويين ذي المسافات والنسبي.

في العادة يواجه الباحثون مواقف تجريبية قد لا تفي بمتطلبات الاختبارات البارامترية، أي عدم استيفائها لشروط الاختبارات المعملية. ففي مثل هذه المواقف، قد لا يكون من الملائم استخدام الاختبارات البارامترية. ومن هنا ينبغي على الباحث أن يضع في اعتباره أنه عندما لم يتم استيفاء شروط استخدام الاختبارات المعملية فإن إجراء الاختبار قد يقود إلى تفسيرات خاطئة للبيانات المستخدمة. من حسن الحظ أنه توجد تقنيات متعددة لاختبار الفرض كبدائل للاختبارات البارامترية. ويطلق على هذه البدائل الاختبارات غير البارامترية. في هذا الفصل سنتناول بعضاً من هذه الاختبارات غير البارامترية شائعة الاستخدام مثل مربع كاي (إحصاء يعرف بمربع كاي لحسن المطابقة ومربع كاي للاستقلالية) ويُسْتَخْدَمُ هذان الاختباران في تقييم الفروض حول التناسب، أو العلاقات الموجودة داخل المجتمعات الإحصائية. إضافة إلى اختبار ثنائي الحد كحالة خاصة لمربع كاي. وتجدر الإشارة إلى أن كلا الاختبارين المتعلقين بمربع كاي مثلهما مثل معظم الاختبارات غير البارامترية لا تتطلب افتراضات أو معلومات حول خصائص التوزيع الأساسي للمجتمع، وربما القيام ببعض الافتراضات (إن وجدت) حول توزيع المجتمع. ولذلك يطلق على الاختبارات غير البارامترية، في بعض الأحيان، اختبارات حرة التوزيع Distribution Free Tests.

إن أكثر الفروض وضوحاً بين الاختبارات البارامترية وغير البارامترية يكمن في نمط البيانات التي تستخدم في كل منهما.

إن كل الاختبارات التي تعاملنا معها حتى الآن، في متن هذا الكتاب، تتطلب درجات عددية Numerical Scores. أما على الجانب الآخر، فإن الاختبارات غير البارامترية، تتعامل مع بيانات.. هذه البيانات ينبغي أن تصنف على هيئة فئات مثل: عال، متوسط، منخفض، عصري، تقليدي... الخ، أي أنه يمكن ملاحظة أن هذه البيانات قد تكون مقاسة على المستوى الاسمي أو الترتيبي، وبالتالي لا تولّد لنا قيماً عدديةً يمكننا استخدامها في حساب المتوسطات الحسابية والتباين. وبدلاً من ذلك، فإن بيانات كثيرة من الاختبارات غير البارامترية، ببساطة بيانات تكرارية.

أحياناً. قد يلجأ الباحث إلى الاختيار بين استخدام الاختبار البارامتري والاختبار غير البارامتري. فاختيار الاختبار غير البارامتري، عادة يستلزم تحويل البيانات من درجات عددية إلى فئات غير عددية. فعلى سبيل المثال، يمكن للباحث أن يبدأ بدرجات عددية لقياس تقدير الذات Self Esteem، وتوليد ثلاث فئات تحتوي على:

درجة عالية لتقدير الذات، درجة متوسطة، ودرجة منخفضة.

وتجدر الإشارة: إلى أنه في معظم الأحوال، يفضل استخدام الاختبار البارامتري باعتباره أكثر ترجيحاً للكشف عن الفرق الحقيقي أو العلاقة الحقيقية. وعلى أية حال، تواجهنا مواقف يكون فيها تحويل الدرجات إلى فئات خياراً أفضل:

1- أحياناً، يكون من البساطة الحصول على مقاييس فئة. على سبيل المثال، من السهولة بمكان تصنيف الطلاب إلى طلاب ذي قدرات عالية، متوسطة أو منخفضة أكثر من الأداء على درجة عددية لقياس قدرة كل طالب على حدة.

2- إن الدرجات الأصلية قد لا تلبي بعض الافتراضات الأساسية لتؤكد إجراءات إحصائية محددة. على سبيل المثال، اختبار t واختبار أنوفا ANOVA يفترضان أن البيانات قد جاءت من توزيعات طبيعية Normal Distributions. كذلك، فإن المقاييس المستقلة للاختبارات هي الأخرى، تفترض أن المجتمعات المختلفة لديها نفس التباين (افتراضية التباين) The Homogeneity of Variance assumption. وعندما يشك الباحث في أن البيانات التي يتعامل معها، بيانات لا تفي بهذه الافتراضات، من هنا يكون من الأفضل تحويل الدرجات إلى فئات، وبالتالي، استخدام التقنيات الإحصائية اللامعلمية لتقييّم بياناته.

3- قد تكون الدرجات الأصلية في الواقع درجات لديها تباين عالٍ. ويعتبر هذا التباين المكون الأساسي للخطأ المعياري Standard Error لمقام معادلة إحصاء t، ومصطلح الخطأ Error Term في مقام معادلة نسبة F. Ratio .F. وعليه، فإن حجم التباين الكبير يمكن أن يقلل، وبشكل كبير، من احتمالية، أن هذه الاختبارات البارامترية سوف تكشف عن فروق دالة. وبتحويل الدرجات إلى فئات سيقلل جوهرياً من

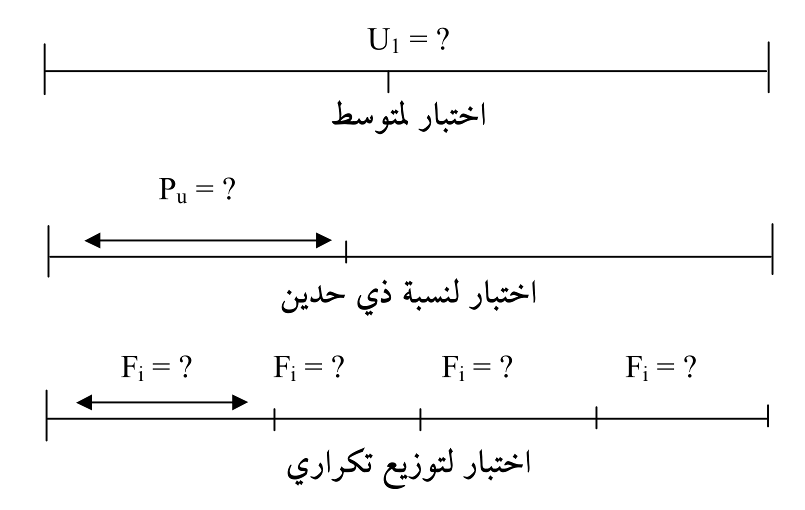
عملية التباين. فعلى سبيل المثال، أن كل الأفراد يتطابقون في الفئات الثلاث (عال، متوسط، منخفض) بغض النظر عن كيفما تكون الدرجات الأصلية[1].

ناقشنا في الفصول السابقة المواقف التي قمنا فيها باختبار فرض حول متوسط مجتمع أو نسبة لتوزيع ذي حدين. كما افترضنا أن متوسط مجتمع أو نسبة بقيمة محددة، وبعدئذ حددت احتمالية سحب العينة من هذا المجتمع بمتوسط أو نسبة تحصلنا عليها حقاً أثناء البحث. ونقوم بفعل ذلك من خلال حساب قيمة Z أو درجة t، ثم نقوم بإيجاد الاحتمالية المقابلة لهذه الدرجات من الجدول المعد لهذا الغرض (جدول توزيعات Z أو جدول توزيعات t). فإذا كان الفرق بين إحصاء العينة Sample Statistic، والقيمة المفترضة للمجتمع فرقاً كبيراً، فالاحتمالية المقابلة لتلك العينة التي تم سحبها من ذلك المجتمع سوف تكون منخفضة. وباختصار، فإن السؤال يختص بما إذا كان الفرق المشاهد بين إحصائية العينة والقيمة المفترضة للمجتمع كبيراً بشكل كافٍ.

اختبار مربع كاي لحسن المطابقة:

يمثل مربع كاي لحسن المطابقة اختبار لا معلمي لتوزيع الحالات عبر مدى واسع من القيم لمتغير واحد. وتستند فلسفة هذا الاختبار على حساب مقياس يعبر عن مدى الفرق بين أعداد المشاهدات والأعداد المتوقع مشاهدتها، فيما إذا كان النموذج الإحصائي صحيحاً. فإذا كان هذا المقياس صغيراً، كان النموذج مقبولاً والعكس بالعكس أي إذا كان هذا المقياس كبيراً فإنه بالتالي يتعذر قبول هذا النموذج الإحصائي[2].

إن طبيعة السؤال المطروح وفقاً لاختبار حسن المطابقة كمقابل لاختبارات أخرى التي تواجه الباحث يمكن توضيحها بالشكل التالي:

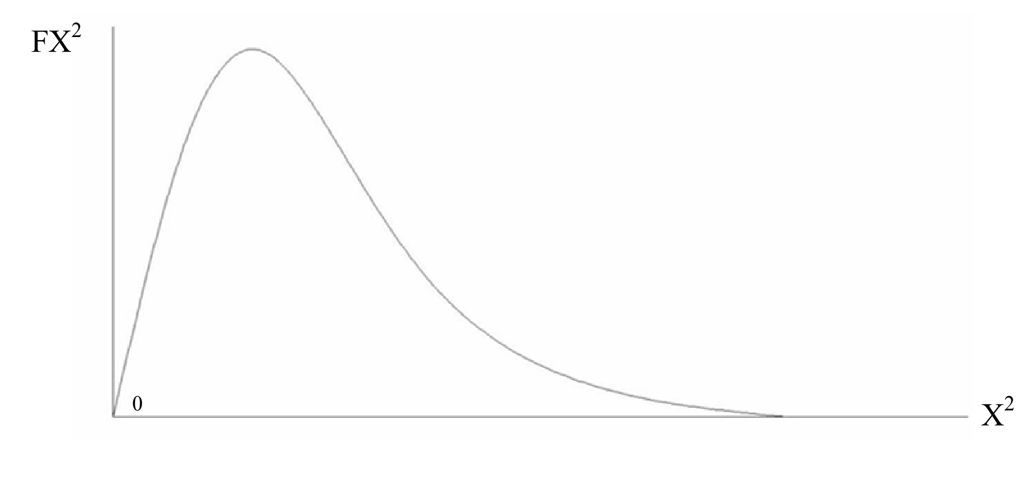


شكل (17 ـ 1) مقارنة بين اختبارات استدلالية

يحلل اختبار مربع كاي لحسن المطابقة التوزيع التكراري. وبما أن التوزيعات التكرارية يمكن بناؤها لبيانات اسمية وبيانات ترتيبية. كما هو الأمر للبيانات ذات المسافات المتساوية والنسبية. (كما بينا ذلك في الفصل الثالث). إذاً، اختبار مربع كاي لحسن المطابقة يمكن أن يستخدم لكل مستويات القياس. في هذا الفصل سنتعامل مع هذا الاختبار كما يطبق للبيانات الاسمية والبيانات الترتيبية حيث تنظم البيانات في فئات منفصلة. وبعد ذلك سوف نبين في سياق هذا الفصل تحليل التوزيع التكراري للبيانات المقاسة على المستويين ذي المسافات المتساوية والنسبي موضحين كيف أن هذا الاختبار يكون اختباراً مفيداً في تحليل التوزيع التكراري.

لقد أُطْلِقَ على هذا الاختبار اختبار مربع كاي لأن توزيع المعاينة الذي يتم استخدامه لتقييم الاحتمالية المتعلقة بالفرض الصفري على أنها صحيحة، هو توزيع مربع كاي (سوف نعود إلى هذه المسألة في الفصل المتعلق باختبار مربع كاي للاستقلال الأكثر استخداماً في مجال البحوث الاجتماعية).

إن الصورة العامة لتوزيع مربع كاي تتضح من خلال الشكل التالي:



شكل (17 ـ 2) توزيع مربع كاي (X^2)

ويتخذ بناء توزيع مربع كاي نفس الأساس كما في توزيعات المعاينة التي تمت مناقشتها. ونعني بتوزيعات المعاينة احتمالية توزيع إحصاء العينة التي نتحصل عليها من عدد لا متناهي من العينات ذات الأحجام المتساوية المسحوبة من مجتمع بخصائص محددة.

ولتوضيح اختبار حسن المطابقة نحاول الإجابة على السؤال التالي:

هل يتأثر معدل الجريمة بفصول السنة؟ عند طرح هذا السؤال فإن الباحث لا يرغب بشكل واضح في معرفة معدل الجريمة بقدر ما يتركز اهتمامه على توزيعات معدلات الجريمة عبر فصول السنة.

الإجراءات:

بادئ ذي بدء ينبغي على الباحث صياغة فرض حول توزيع المجتمع: يفترض أنه لا توجد علاقة بين معدلات الجريمة وفصول السنة. ومن خلال هذا الفرض يتوقع الباحث أن عدد الجرائم التي ارتكبت في أي عام تُوَزَّع بشكل متساوٍ عبر الفصول الأربعة.

$$Fe = \frac{\text{العدد الإجمالي للجرائم (N)}}{4}$$

حيث إن Fe تعني التوزيعات التكرارية في كل فئة.

من ناحية ثانية، يمكننا القول، بأن معدل الجريمة في أي سنة قد يتأثر بأحداث عشوائية تسبب فرقاً طفيفاً في التوزيع عن تلك النتيجة المتوقعة. بمعنى آخر، أنه ليست كل عينة من العينات ستتطابق مع هذا التوقع لعدد متساوٍ من الجرائم المرتكبة في كل فصل من فصول السنة. ويمكننا توضيح الفرق بين القيمة المتوقعة والقيمة المشاهدة بحساب إحصاء عينة كاي تربيع:

$$x^2 = \sum \frac{(Fo - Fe)^2}{Fe} \quad \text{العينة}$$

حيث إن: Fe : التوزيعات المتوقعة في كل فئة.

F0 : التوزيعات المشاهدة في كل فئة.

تجدر الإشارة إلى القول بأنه إذا كانت نتيجة العينة تطابق تماماً النتيجة المتوقعة، فإن قيمة مربع كاي للعينة تطابق تماماً النتيجة المتوقعة، فإن قيمة مربع كاي للعينة ستكون صفراً(0).

إن السؤال الذي يمكن طرحه هو: ما هو الموقف عندما يكون التوزيع المشاهد ليس على نحو دقيق مساوياً للتوزيع المتوقع؟

بالنظر إلى معادلة X^2 يمكننا مشاهدة أن أي فرق سيولد قيمة إيجابية لعينة مربع كاي. وهذا سببه أن أي فرق قد تم تربيعه. وبذلك يمكننا التخلص من القيم السالبة. كذلك يمكننا ملاحظة أنه كلما كان الفرق كبيراً بين التوزيعات المشاهدة والتوزيعات المتوقعة (البواقي Residuals)، كانت القيمة عالية (موجبة) لعينة مربع كاي. إن السؤال إذاً يصبح حول ما هي النقطة الأساسية وهي هل القيمة المتعلقة بعينة مربع كاي تصبح كبيرة جداً إلى درجة أن هذه القيمة تقترح أن العينة لم يتم اختيارها من مجتمع تنتشر فيه معدلات الجريمة بشكل متسق عبر فصول السنة؟

يمكننا اختيار القيمة الحرجة لـِ X^2 بنفس الطريقة المتبعة في الاختبارات التي تناولناها في الفصول السابقة. وكذلك يمكننا اختيار مستوى الفا (α) مثل 0.05، كما يمكننا

استخدام جدول توزيع قيم مربع كاي الحرجة (انظر الملحق) لإيجاد القيمة الحرجة لمربع كاي بمستوى معين من الفا آخذين في الاعتبار عدد درجات الحرية (df) لأي توزيع معطى، وللحصول على درجة الحرية تستخدم المعادلة التالية:

$$df = k - 1$$

حيث إن: K عدد الفئات

وعليه، إذا كان لدينا متغير يتألف من أربع فئات كما هو الحال في المثال الذي نتعامل معه (معدلات الجريمة حسب فصول السنة)، فإن درجة الحرية ستكون كالتالي:

$$df = 4 - 1 = 3$$

ومن خلال مراجعة الجدول المتعلق بتوزيعات القيم الحرجة لمربع كاي بمستوى الفا 0.05 بدرجة حرية 3، فإن قيمة مربع كاي الحرجة تساوي

$$x^2 = 7.815(\alpha = 0.05, df = 3)7.815$$

الآن، يمكننا حساب قيمة العينة لمربع كاي (X^2) لنرى ما إذا كانت قيمة العينة تقع داخل المنطقة الحرجة (منطقة الرفض)، على سبيل المثال، إذا لاحظنا فعلياً توزيع (الافتراضي) الجريمة كما هو مبين في الجدول (1). فالسؤال هو: هل يمكننا أن نخلص إلى القول بأن الجريمة متأثرة بفصول السنة؟

جدول (17 ـ 1) توزيع الجريمة حسب فصول السنة

	الصيف	الربيع	الشتاء	الخريف	المجموع
المشاهد	300	270	200	250	1020
المتوقع	255	255	255	255	1020
البواقي	45	15	−55	−5	

المصدر: .George Argyrous, Statistics for Social and Healh Research, op. Cit , P.325

للحصول على القيم المتوقعة يمكننا استخدام المعادلة التالية:

$\frac{1020}{4} = 255$ = تقسيم المجموع الكلي على عدد الفصول.

إن الصف الذي يطلق عليه البواقي في الجدول يعني الفرق بين القيم المشاهدة والقيم المتوقعة.

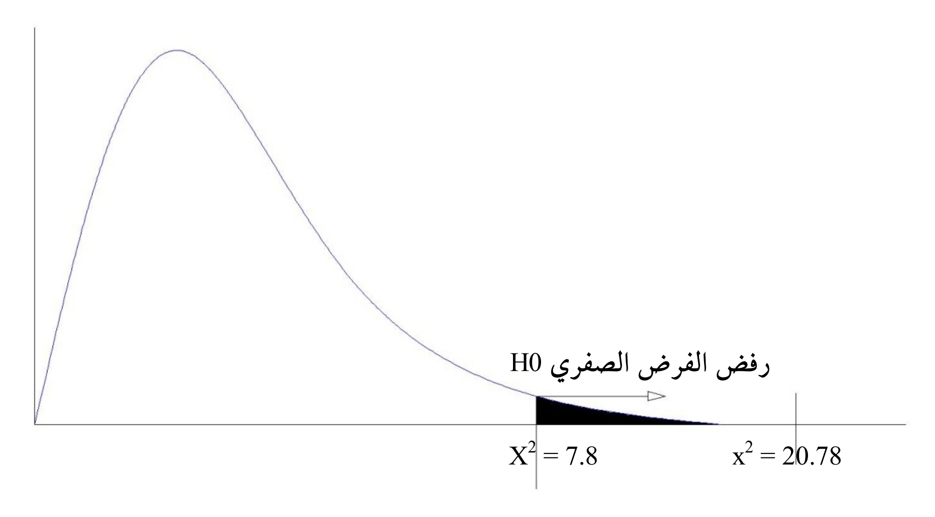
الآن يمكننا أن نستعيض عن نتائج العينة إلى المعادلة المتعلقة بمربع كاي:

العينة $X^2 = \sum \frac{(Fo - Fe)^2}{Fe}$

$$= \frac{(300-255)^2}{255} + \frac{(270-255)^2}{255} + \frac{(200-255)^2}{255} + \frac{(250-255)^2}{255}$$
$$= 20.78$$

والشكل التالي يوضح درجات العينة والقيم الحرجة X^2 لمساعدتنا في اتخاذ القرار.

ومن خلال هذا الشكل، تبين لنا أن قيمة مربع كاي للعينة بشكل واضح تقع داخل المنطقة الحرجة، الأمر الذي يقودنا إلى رفض الفرض الصفري للتوزيع المتساوي للجريمة عبر فصول السنة [3].



شكل (17 ـ 3) توزيع (X^2) ودرجة الحرية (df = 3)

إجراء اختبار مربع كاي لحسن المطابقة باستخدام Spss :

1- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Analyze ⟸ Non Parametric tests ⟸ Chi - Square

2- انقر على Crime by Season في القائمة.

3- انقر على ◄ يقوم بلصق Crime by Season في Test Variable List.

4- انقر على ok.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

Chi - Square Test

Frequencies

	Observed (N)	Expected (N)	Residual
Summer	300	255.0	45.0
Spring	270	255.0	15.0
Autumn	200	255.0	-55.0
Winter	250	255.0	-5
Total	1020		

Test Statistics

	Crime by Season
Chi - Square[a]	20.784
df	3
Asymp. sig	000

a. o cells (o %) have expected frequencies less than 5
The Minumum expected cell frequencies is 255.0

المصدر: .George Argyrous, Statistics for Socialand Healh Research, op. Cit , P.328

شكل (17 ـ 4) مخرجات Spss لاختبار مربع كاي

تفسير مخرجات SPSS لمربع كاي (X^2) :

يحتوي المربع المُعَنْوَن: الجريمة وفقاً لفصول السنة Crime by Season على الإحصاء الوصفي الذي يلخص العينة المدروسة. وفي هذه الحالة فإن توزيع الحالات عبر الفصول الأربعة من فصول السنة قد يوفرها عمود التوزيعات المتوقعة التي تولد لنا العدد المتساوي لكل الحالات المتوقعة في كل فصل من فصول السنة.

إن القيم المتوقعة لعمود (N) قد طرحت من عمود (N) القيم المشاهدة تعطينا الفرق بين القيم المشاهدة والقيم المتوقعة.

أما المربع الثاني المُعَنْوَنُ بـِ إحصاء الاختبار لمربع كاي Chi - Square Test Statistics، نجد أن قيمة مربع كاي تساوي 20.784 بدرجة حرية 3 df(3) بدلالة تصل إلى 1 في كل 10.000 عينة إذا ما وزعت الجريمة بشكل متساوٍ عبر فصول السنة. إن مثل هذه الاحتمالية المنخفضة تقودنا إلى رفض الفرض الصفري: معدلات الجريمة يبدو أنها مرتبطة بفصول السنة[4].

اختبار مربع كاي لحسن المطابقة للحالة السوية (الاستواء Normality) :

من خلال المثال الذي تعاملنا معه في اختبار مربع كاي لحسن المطابقة للبيانات المقاسة على المستوى الاسمي التي تقع ضمن فئات منفصلة. وفي هذا السياق يمكننا القول بأن أي اختبار يمكن تطبيقه على بيانات اسمية وبيانات ترتيبية مع ذلك، يمكن تطبيقه على مستويات أعلى من المقياس كالبيانات ذات المسافات المتساوية والنسبية. ففي حالة المقياس ذي المسافات والنسبي فإننا ننظر إلى التوزيع التكراري للحالات عبر مدى من القيم أو عبر فئات متساوية تماماً بنفس الطريقة التي ينظر فيها إلى التوزيع عبر فئات منفصلة.

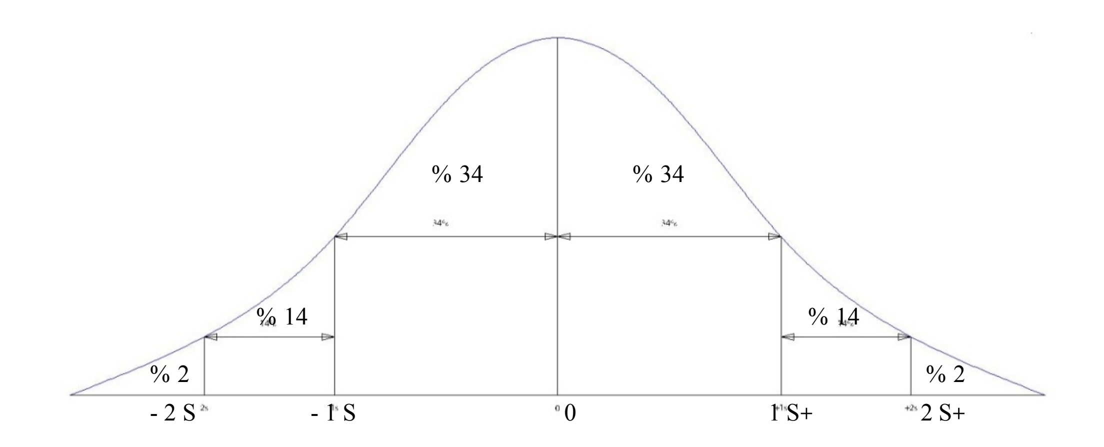
إن هذا هو المنطق الذي يجعل من اختبار حسن المطابقة مقياساً مفيداً في تقييم ما إذا كانت البيانات المقاسة على المستويين ذي المسافات والنسبي جاءت من مجتمع طبيعي. من هذه الناحية، فإن هذا الاختبار هو اختبار لا معلمي Non - Parametric Test يمكن أن يكون مفيداً كاختبار تمهيدي وتكملة للاختبارات المعلمية Parametric Tests التي تتطلب الافتراض أن العينة سحبت من مجتمع طبيعي [5].

مثال: نفترض أن لدى الباحث عينة ويرغب في تقييم ما إذا كان قد تم سحبها من مجتمع طبيعي. وكما سبقت الإشارة، فقد عرف التوزيع الطبيعي بالتوزيعات التكرارية كما هو مبين في الجدول رقم (2) وشكل رقم (5).

جدول (17ـ2) توزيع المنحنى الطبيعي

مدى القيم Range of Value	نسبة الحالات Percentage of Cases
أبعد من 2 انحراف معياري تحت المتوسط	% 2
بين 1 و 2 انحرافات معيارية تحت المتوسط	% 14
داخل 1 انحراف معياري تحت المتوسط	% 34
داخل 1 انحراف معياري فوق المتوسط	% 34
بين 1 و 2 انحرافات معيارية فوق المتوسط	% 14
أبعد من 2 انحرافات معيارية فوق المتوسط	% 2

المصدر: .George Argyrous, Statistics for Social and Healh Research, op. Cit , P.331



شكل (17ـ5) المساحات تحت المنحنى الطبيعي

إنه بإمكاننا استخدام القيم النسبية كما هو مبين في الجدول رقم (17 - 2) لحساب القيم المتوقعة باستخدام المعادلة المتعلقة بمربع كاي. لاحظ مرة أخرى أن هذا المثال لا يشبه المثال السابق حول معدل الجريمة حسب فصول السنة، حيث إننا لا نفترض هنا أن الحالات موزعة توزيعاً متساوياً عبر الفئات، وإنما بدلاً من ذلك فالتوزيعات المتوقعة قد استندت على خصائص المنحنى الطبيعي.

مثال[6]:

نفترض أن لدى الباحث عينة مؤلفة من 110 أشخاص بمتوسط عمري يصل إلى 45 سنة، وبانحراف معياري 10 سنوات، إذا كانت هذه العينة موزعة توزيعاً طبيعياً، فالباحث يتوقع أن يجد عدد الأشخاص داخل المدى المبين الذي يبينه الجدول رقم (17 - 3).

جدول (17 ـ 3) التوزيع المتوقع للعينة

مدى القيم	نسبة الحالات	عدد الحالات
25 سنة أو أقل (أبعد من 2 انحرافات معيارية تحت المتوسط)	% 2	$2.2 = 110 \times \frac{2}{100}$
26 - 35 سنة (بين 1 و 2 انحرافات معيارية تحت المتوسط)	% 14	$15.4 = 110 \times \frac{14}{100}$
36 - 45 سنة (داخل 1 انحراف معياري تحت المتوسط)	% 34	$37.4 = 110 \times \frac{34}{100}$
46 - 55 سنة (داخل 1 انحراف معياري فوق المتوسط)	% 34	$37.4 = 110 \times \frac{34}{100}$
56 - 65 سنة (بين 1 و 2 انحراف معيار فوق المتوسط)	% 14	$15.4 = 110 \times \frac{14}{100}$
66 سنة أو أكبر (أبعد من 2 انحراف معياري فوق المتوسط)	% 2	$2.2 = 110 \times \frac{2}{100}$

من ناحية ثانية، يمكننا فعلياً الحصول على توزيعات عينة كما هو مبين في الجدول التالي:

جدول (17 ـ 4) التوزيعات المشاهدة للعينة

مدى القيم	عدد الحالات
25 سنة أو أقل	5
26 - 35 سنة	17
36 - 45 سنة	33
46 - 55 سنة	33
56 - 65 سنة	17
66 سنة أو أكبر	5

من خلال المقارنة بين القيم المشاهدة والقيم المتوقعة يتجلى لنا بوضوح الفرق بين القيم المشاهدة والقيم المتوقعة: والسؤال الذي يُطْرَح في هذا السياق هو: هل هذا الفرق يقودنا إلى رفض الفرض الذي مفاده أن المجتمع موزع توزيعاً طبيعياً؟

للإجابة على هذا السؤال، يتطلب منا حساب قيمة مربع كاي(X^2):

$$x^2 = \sum \frac{(Fo - Fe)^2}{Fe} \quad \text{العينة}$$

$$= \frac{(5-2.2)^2}{2.2} + \frac{(17-15.4)^2}{15.4} + \frac{(33-37.4)^2}{37.4}$$
$$+ \frac{(33-37.4)^2}{37.4} + \frac{(17-15.4)^2}{15.4} + \frac{(5-2.2)^2}{2.2}$$
$$= 8.5$$

بعد أن أجرينا إحصاء اختبار مربع كاي، ينبغي علينا مقارنة درجة العينة مع القيمة الحرجة لمربع كاي التي تستند في الأساس على درجة الحرية (df). في هذا المثال، لدينا

ست فئات، وبالتالي فإن درجة الحرية تساوي 5 = 1 – 6، على مستوى ألفا (α) 05.، والقيمة الحرجة لمربع كاي هي 11.070 (انظر ملحق توزيع القيم الحرجة لمربع كاي).

ولما كانت القيمة المحسوبة لمربع كاي أصغر من القيمة الحرجة، عليه، لا يمكننا رفض الفرض الصفري الذي مفاده أن العينة جاءت من مجتمع موزع توزيعاً طبيعياً.

خلاصة:

لقد تم مناقشة اختبار X^2 لحسن المطابقة. وبالرغم من أن اختبار X^2 مرتبط بحسابات مختلفة قليلاً، إلا أنه يشبه إلى حد كبير الاختبار "ذا الحدين" الذي سنتناوله لاحقاً. حيث أن اختبار Z لتوزيع ذي الحدين للنسب يطبق فقط لتوزيعات تكرارية تم تنظيمها في توزيع ذي حدين.

إن اختبار مربع كاي يكون اختباراً أكثر شمولاً عند تطبيقه على التوزيعات التكرارية لأي عدد من الفئات (وبالتالي فإن اختبار Z لتوزيع ذي حدين للنسب يمكن اعتباره حالة خاصة لاختبار X^2.

إن هذا الوضع يعطي اختبار X^2 تطبيقات واسعة إلى عينتين أو أكثر كما سنرى في الفصل المتعلق باختبار مربع كاي للاستقلال.

أسئلة للمراجعة:

1- ما هي درجات الحرية والقيم الحرجة لـِ X^2 لكل من:

الفا 0.10 = α و الفا 0.05 = α لاختبار حسن المطابقة لمتغير يحتوي على:

أ- ثلاث فئات ب- خمس فئات ج- ثمان فئات.

2- احسب اختبار حسن المطابقة للبيانات التالية لاختبار الفرض الذي مفاده: أن العينة التي تم سحبها من مجتمع ذي توزيع متماثل لحالات عبر الفئات:

أ-

القيم	عدد الحالات
1	45
2	40
3	55
4	54
5	38

ب-

القيمة	عدد الحالات
1	120
2	111
3	119
4	125
5	120
6	127
7	118

3- تم مسح 90 شخصاً لمعرفة كمية الوقت الذي يتم استغلاله في القراء كل يوم، وتم قياس كمية الوقت. يرغب الباحث في اختبار الفرض الذي مفاده أن العينة التي تم سحبها من مجتمع موزع توزيعاً طبيعياً. علماً بأن متوسط القراءة للعينة المسحوبة

يصل إلى 45 دقيقة، بانحراف معياري 15 دقيقة. وأن توزيع هذه القيم المشاهدة تظهر بالشكل التالي:

مدى القيم	عدد الحالات
أقل من 16 دقيقة	3
16 – 30 دقيقة	15
31 – 45 دقيقة	34
46 – 60 دقيقة	31
61 – 75 دقيقة	5
أكثر من 75 دقيقة	2

- مستخدماً ألفا 0.05 = α لاختبار فرضية الاستواء (الحالة السوية) للمجتمع، أجر اختبار حسن المطابقة؟

4- تم مقارنة خمس مدارس فيما يتعلق بنسبة الطلاب الذين يتابعون تعليمهم الجامعي. عينة من 50 طالباً قد تخرجوا من كل مدرسة تم سحبهم. وعدد أولئك الذين دخلوا الجامعة من كل مدرسة هم:

المدرسة	العدد الذي دخل الجامعة
مدرسة 1	22
مدرسة 2	25
مدرسة 3	26
مدرسة 4	28
مدرسة 5	33

المطلوب:

أ- حساب القيم المتوقعة، وبعد ذلك حساب قيمة X^2 لاختبار حسن المطابقة.

ب- ما هي النتائج التي توصلت إليها حول توقع دخول الجامعة من كل مدرسة من هذه المدارس؟

الهوامش والمصادر:

أولا: الهوامش:

1. Frederick J. Gravetter and Larry B. Wallnau , Statistics for The Behavioral Sciencies , Wadsworth Cengage Learning , USA , 2010 , PP. 606 - 607.
2. George Argyrous, Statistics for Social & Health Research, with Aguide to Spss , SAGE Publications, London, 2001 , P. 323.
3. Ibid., P. 327.
4. Ibid., P. 331.
5. Ibid., P. 331.
6. Ibid., PP. 332 - 334.

ثانيا: المصادر:

1- Frederick J. Gravetter and Larry B. Wallnau , Statistics for The Behavioral Sciencies , Wadsworth Cengage Learning , USA , 2010.

2- George Argyrous, Statistics for Social & Health Research, with aguide to Spss , SAGE Publications, London, 2001.

3- Hugh Coolicam , Hodder Arnold , Research Methods and Statistics in Psychology , 4th ed , Hodder Stoughton Educational , London , 2004.

الفصل الثامن عشر

اختبار مربع كاي (X^2) للاستقلال

مقدمة:

نتناول في هذا الفصل أحد التقنيات الإحصائية التي يمكن من خلالها إجراء اختبار فرضية لبيانات مقولية تم ترتيبها في جداول متقاطعة. ويطلق على هذه التقنية الإحصائية اختبار مربع كاي (X^2) للاستقلال. ويشبه اختبار مربع كاي للاستقلال اختبار عينة واحدة التي تم تناولها في موضع آخر من هذا الكتاب.

اختبار مربع كاي واختبارات الدلالة الأخرى:

في الفصول السابقة لهذا الكتاب تم التركيز على اختبار الاستدلال الذي تم تحديده من خلال معيارين أساسيين أولهما: الإحصاء الوصفي المستخدم في وصف البيانات الخام لعينة. وثانيهما: عدد العينات التي تم وصفها من خلال الاستدلال الذي تم القيام به.

1- الإحصاء الوصفي المستخدم في وصف البيانات الخام:

إن هذا العامل في حد ذاته دالة لقضيتين أساسيتين: أولهما: متعلقة بالسؤال البحثي المطروح الذي يتعين على الباحث الإجابة عليه. إن السؤال البحثي تقريباً وبشكل ثابت يوجه اهتمامات الباحث نحو خصائص محددة لتوزيع متعلق بالمتغير الخاضع للاستقصاء. على سبيل المثال، أن الباحث في شئون الصحة العامة، قد يَنْصَبُّ اهتمامُه حول سؤال مفاده ما إذا كان مجتمع ما يكون فيه المتوسط العمري يميل نحو الشباب أو الكبار. إذاً مشكلة البحث هذه تقود الباحث إلى استخدام إحصاءات النزعة المركزية لهذا المتغير. وكذلك الأمر بالنسبة للباحث في العلوم السياسية الذي قد ينصب اهتمامه أيضاً على معرفة التوزيع العمري لهذا المجتمع، إلا أنه في هذه الحالة فإن التركيز يمكن أن ينصب على العينة العمرية التي يحق لها المشاركة في العملية السياسية. من هنا نجد أن الباحث سيقوم بتنظيم بياناته من خلال توزيع ذي حدين وحساب نسبة العينة الأعلى والأدنى للفئة العمرية. إن كلا الباحثين يرغبان في التعامل مع نفس المجتمع، وإن كليهما، لديه نفس البيانات الخام. إلا أن السؤال البحثي المطروح هو الذي يحدد ما إذا كان الباحثان يرغبان في تبني توزيعات النزعة المركزية أو نسب الحالات الأعلى والأدنى من نقطة محددة على المقياس. وثانيهما يتعلق بالقضية التي تحدد الإحصاء الوصفي الذي يستخدم لتلخيص البيانات وهي مستوى القياس للمتغير. إن مستوى القياس كثيراً ما يقيد الطريقة التي يمكن أن نصف بها خصائص التوزيع الذي يرغب الباحث فيه. على سبيل المثال، نفترض أننا نرغب في معرفة النزعة المركزية للتوزيع العمري لعينة ما، وسواء قمنا بجمع بيانات مقاسة على المستوى ذي المسافات والنسبي، أو مقاسة على المستوى الترتيبي فهي التي تحدد مقاييس النزعة المركزية (المنوال، الوسيط، أو المتوسط) فنحن في الواقع نستخدم هذه الإحصاءات الوصفية لتلخيص البيانات الخام. فإذا تم قياس العمر بالسنوات (مقياس ذو المسافات المتساوية والنسبي)، فإنه في هذه الحالة يعتبر الإحصاء الوصفي المستخدم لتلخيص البيانات هو المتوسط الحسابي. أما إذا تم تصنيف العمر حسب الفئات في مدى يتراوح من: صغير جداً إلى كبير جداً (مقياس ترتيبي) إذاً في هذه الحالة فإن الإحصاء الوصفي المستخدم هو الوسيط. إن هذين العاملين المتمثلين في

السؤال البحثي ومستوى القياس معاً يحددان الإحصاء الوصفي الفعلي الذي على ضوئه يتم إجراء الاختبارات الاستدلالية.

عدد العينات التي ينبغي مقارنتها:

كما رأينا في الفصل المتعلق باختبار t لمتوسط حسابي لعينة واحدة، فإنه عندما يتم جمع البيانات من عينة واحدة فقط، حينئذ يكون لدينا مدى معين لاختبارات الاستدلال للاختيار من بينها: أنه يمكن القول بأن مدى هذا الاختيار يختلف عند التعامل مع عينتين، ولذلك ينبغي علينا في هذه الحالة أن نقوم باستدلال حول كل واحد من هذين المجتمعين اللذين سُحِبَتْ منهما هاتان العينتان. وبالتماثل، إذا كانت لدينا أكثر من عينتين، فإننا في هذه الحالة نواجه مدى آخر من الاختبارات للاختيار من بينها. على سبيل المثال، عند مقارنة متوسطات باختبار t لمتوسطات العينة يكون ملائماً لعينة أو عينتين. في حين أن تحليل ANOVA يستخدم لأكثر من عينتين وهكذا. واضعين هذه القضايا بعين الاعتبار، فإنه بإمكاننا النظر إلى الأحوال التي يكون فيها استخدام مربع كاي ملائماً.

إن الإحصاء الوصفي الذي على ضوئه يستخدم مربع كاي للاستقلال هو التوزيع التكراري المتضمن في جدول ثنائي bivariate table. في الفصل السابع من هذا الكتاب قد بيّنا عملية بناء واستخدام الجداول الثنائية أي جداول التقاطع. وتعد جداول التقاطع جداول ملائمة في تلخيص وعرض البيانات المقولية عند الرغبة في التوزيع الكلي للحالات عبر مجموعة المدى الكلي للفئات مفضلاً ذلك أو بدلاً من مقاييس النزعة المركزية. وتعتبر جداول التوافق ذات أهمية كبيرة وذلك لشيوع استخدامها في وصف البيانات المقاسة على المستوى الاسمي والترتيبي.

وتجدر الإشارة إلى أن البيانات المقاسة على المستويين ذي المسافات والنسبي يمكن اختزالها إلى فئات منفصلة مثل اختزال الدخل إلى فئات: ذوي الدخل المنخفض والمتوسط، والعالي. كما تعتبر جداول التقاطع أداة مهمة لوصف البيانات ذات المسافات المتساوية والنسبية كما هو الحال في أهمية البيانات الاسمية والترتيبية.

إن اختبار مربع كاي يعتبر في الأساس الشيء نفسه بغض النظر عما إذا كان لدينا

عينة واحدة، أو عينتان أو أكثر من ذلك. وكما بينا في الفصل المتعلق باختبار مربع كاي لعينة واحدة لتوزيع تكراري. إن مربع كاي، ليس كالاختبارات الأخرى، فهو اختبار يمكن توسيعه ليشمل عينتين أو أكثر من عينتين من الحالات دون تعديلات كبيرة: فالباحث يمكنه اتباع الإجراء نفسه، والمعادلة نفسها، بغض النظر عن عدد العينات الخاضعة للمقارنة (لاحظ أنه في حالة العينة الواحدة، فقد أطلق على مربع كاي اختبار حسن المطابقة A goodness-of-fit-test في حين يطلق على مربع كاي للاستقلال A test for Ivdependence، عند التعامل مع عينتين أو أكثر).

الاستقلال الإحصائي Statistical Independence:

يمدنا مفهوم الاستقلال الإحصائي بطريقة أخرى تمكننا من معرفة الفرق بين الفرض البحثي والفرض الصفري عندما يكون لدينا متغيران مستقلان إحصائياً فإن أي تغير في واحد من المتغيرين ليس له علاقة بالتغير في المتغير الآخر لأنهما يختلفان بشكل مستقل عن بعضهما البعض. أما على الجانب الآخر، إذا كان هذان المتغيران يعتمدان على بعضهما البعض فإن أي تغير في أحدهما يرتبط بالتغير في المتغير الآخر.

مثال:

دعنا نعيد العمل من خلال المثال الذي أوردناه في الفصل السابق لبناء جدول تقاطعي بين متغيري (الدخل ومكان الإقامة).

جدول (18 ـ 1) العلاقة بين الدخل ومكان الإقامة

مكان الإقامة	الدخل: منخفض	الدخل: عال	المجموع
ريفي	205 (%55)	118 (%34)	323 (%45)
حضري	167 (%45)	230 (%66)	397 (%55)
المجموع	372 (%100)	348 (%100)	720 (%100)

المصدر: George Argyrous, Statistics for Social & Health Research, op.cit, p.395.

بالنظر إلى التوزيعات النسبية لكل خلية من خلايا الجدول يمكننا تقييم ما إذا كان في العينة أن المتغيرين مستقلان، أو ما إذا كان في الحقيقة أنه يُوجد نوع من العلاقة. ومن خلايا بيانات هذا الجدول نلاحظ وجود بعض العلاقة (أو نمط الاعتماد Pattern of Independence) بين هذين المتغيرين. ويبدو من خلال الأرقام المبينة في الجدول أعلاه، أن الأفراد ذوي الدخول المنخفضة يتجهون إلى العيش في المناطق الريفية. في حين أن الأفراد ذوي الدخول المرتفعة يميلون إلى العيش في المناطق الحضرية.

إضافة إلى جدول التقاطع. فقد تم حساب مقاييس التطابق الملائمة لتكميم هذه العلاقة المشاهدة بين هذين المتغيرين. ولما كان على الأقل أحد هذين المتغيرين تم قياسه على المستوى الاسمي، فإن المقياس المناسب هو لامبيدا Lambda وأن قيمة لامبيدا لهذا الجدول تساوي 0.12 وتشير هذه الدرجة إلى علاقة ضعيفة لهذه البيانات.

ومن ناحية ثانية، فإن النتيجة التي توصلنا إليها قد اعتمدت على بيانات عينة، وعليه ينبغي علينا أن نكون حذرين من أن هذه النتيجة ربما قد جاءت نتيجة للتباين العشوائي عند عملية المعاينة من المجتمعات التي لا توجد فيها أي علاقة بين الدخل ومكان الإقامة. وأن اختبار مربع كاي كفيل بتقييم هذه الاحتمالية.

حساب وتفسير إحصاء الاختبار لجداول مربع كاي للاستقلال:

يمتلك مربع كاي عنصرين أساسيين هما إحصاء الاختبار الذي يطلق عليه القيمة المتحصل عليها لمربع كاي. والعنصر الثاني القيمة الحرجة للإحصاء التي تمثل القيمة لإحصاء الاختبار الذي ينبغي أن يمتد لرفض الفرض الصفري.

إن إحصاء الاختبار لمربع كاي ليس مقياساً للتطابق لأن قيمة مربع كاي تكون متأثرة بالدرجة التي يكون فيها المتغيران معتمدين على بعضهما البعض. إن قيمة مربع كاي سوف تكون أكبر إذا كان المتغير التابع يعتمد على المتغير المستقل. في حين تكون قيمة مربع كاي أصغر إذا كان المتغير التابع يختلف بشكل مستقل بدون النظر إلى المتغير المستقل[1].

إن أول خطوة لإجراء مربع كاي للاستقلال، كما هو الحال، في كل مقاييس الاستدلال هي البدء بصياغة الفرض الصفري والفرض البديل، ففي المثال الذي أمامنا، فإن صياغة الفرض الصفري تتخذ الشكل التالي:

Ho: الدخل ومكان الإقامة متغيران مستقلان عن بعضهما البعض.

Hi : الدخل ومكان الإقامة ليسا متغيرَيْن مستقلين عن بعضهما البعض.

إذا رفضنا الفرض الصفري الذي مفاده أن المتغيرين مستقلان، فإننا في هذه الحالة نستنتج أن المتغيرين ليسا مستقلين في المجتمع المدروس. وعلى العكس من ذلك، إذا عجزنا عن رفض الفرض الصفري، فإننا نجادل في أن المتغيرين حقاً هما مستقلان بالرغم من أن التبعية قد تم ملاحظتها في العينات.

إذا نظرنا إلى مثالنا الحقيقي، فإننا نكون قد حددنا أن المتغيرين ليسا مستقلين في العينة - أي لا تبدو أن هناك بعض العلاقة - ولكن هل باستطاعتنا استخلاص هذا الاستدلال حول المجتمعات التي سحبت منها هذه العينات؟.

ولكي نرى كيف يساعدنا اختبار مربع كاي في تقييم ما إذا كان هذان المتغيران حقاً مستقلين عن بعضهما البعض بالرغم من وجود تبعية في العينات. دعنا نبدأ بالنظر إلى البيانات الخام والتوزيعات النسبية في الجدول التالي:

جدول (18 ـ 2) مكان الإقامة: كل المبحوثين

مكان الإقامة	المجموع	النسبة المئوية
ريفي	323	% 45
حضري	397	% 55
المجموع	720	% 100

إن مجاميع الصف والنسب المئوية هي النقاط المرجعية التي من خلالها يتم إجراء اختبار مربع كاي. إن الحجة تكون أنه إذا كان 45% من مجموع المبحوثين يعيشون في

مناطق ريفية، حينها نتوقع أن 45% من كل مجموعة (دخل منخفض، ودخل عالٍ) هم أيضاً يعيشون في مناطق ريفية إذا كان المتغيران مستقلين.

الجدول التالي يوضح هذه التوزيعات النسبية المتوقعة.

جدول (18ـ3) التوزيعات النسبية المتوقعة للخلية

	الدخل		
مكان الإقامة	منخفض	عال	المجموع
ريفي	% 45	% 45	% 45
حضري	% 55	% 55	% 55
المجموع	% 100	% 100	% 100

من ناحية ثانية، حتى ولو أن الفرضية الصفرية للاستقلال صحيحة، فإنه ليس في كل الأحوال أن نتوقع دائماً عينات عشوائية لأولئك الذين يكسبون دخولاً منخفضة ودخولاً عالية لتعكس ذلك. فعلى سبيل المثال، ربما أحياناً قد نسحب عينات من مجموعات ذوي دخول منخفضة وعالية ونحصل على واحدة من النتائج الثلاث التالية كما هو موضح في الجدول التالي:

جدول (18 – 4 أ) عينة 1

	الدخل		
مكان الإقامة	منخفض	عال	المجموع
ريفي	% 47	% 44	% 45
حضري	% 53	% 56	% 55
المجموع	% 100	% 100	% 100

جدول (18 – 4 ب) عينة 2

	الدخل		
مكان الإقامة	منخفض	عال	المجموع
ريفي	% 52	% 40	% 45
حضري	% 48	% 60	% 55
المجموع	% 100	% 100	% 100

جدول (18 – 4 ج) عينة 3

	الدخل		
مكان الإقامة	منخفض	عال	المجموع
ريفي	% 65	% 30	% 45
حضري	% 35	% 70	% 55
المجموع	% 100	% 100	% 100

المصدر: George Argyrous, Statistics for Social & Health Research, op.cit , p.397.

العينة 1: تصور لنا هذه العينة موقفاً تعكس فيه النسب المشاهدة النسب المتوقعة Expected Percentages، إلى حد بعيد مفترضين أن المتغيرين مستقلان. وفي أحيان أخرى، ربما نجد موقفاً آخر يظهر في العينة الثانية حيث يمكننا ملاحظة وجود تباين كبير بين المجموعات، ولكن هذا التباين ليس بالحد الكبير جداً. في حين تعكس العينة 3: موقفاً متطرفاً نتوقع من خلاله الحصول على حالات من أي طرف من المقياس مسببة التوزيعات النسبية في أول العمودين لتنحرف كثيراً عن تلك التوزيعات النسبية لمجموع العمود. وبالرغم من أن هذه احتمالية عند المعاينة العشوائية من مجتمعات حيث لا توجد علاقة. وهي أيضاً بعيدة الاحتمال.

في الحقيقة أنه باستطاعتنا أن نأخذ عدداً لا متناهياً من العينات العشوائية من مجتمعات يكون فيها المتغيران مستقلين ومشاهدة توزيع النتائج.

وبوضوح فإن معظم النتائج ستكون مشابهة لما لاحظناه في العينتين 1 و 2، وأن قليلاً من النتائج التي تشبه نتائج العينة رقم 3. أن إحصاء مربع كاي يعتبر أداة فعالة، تساعد الباحث في معرفة هذا الفرق بين التكرارات المشاهدة والتكرارات المتوقعة.

ويحسب إحصاء مربع كاي من الفرق بين التكرارات المشاهدة والتكرارات المتوقعة لكل خلية من خلايا الجدول الثنائي Bivariate Table، وأن المعادلة المستخدمة في حساب مربع كاي هي

$$X^2 = \sum\frac{(fo-fe)^2}{fe}$$

حيث إن: F0 تشير للتوزيعات المشاهدة في كل خلية.

Fe: تشير للتوزيعات المتوقعة في كل خلية.

أحياناً يقوم الباحث بسحب عينات تكون "حقيقية" للمجتمع، وبالتالي لا يوجد فرق بين التكرارات الفعلية، والتكرارات المتوقعة. بمعنى آخر، يمكن للباحث أن يتحصل على توزيعات خلية مماثلة لتلك التوزيعات في الجدول رقم 3. في هذه الحالة يتحصل الباحث على قيمة مساوية لصفر لمربع كاي:

$$fo = fe \rightarrow (fo-fe)^2 = 0 \rightarrow X^2 = 0$$

وتجدر الإشارة هنا إلى أنه لن يكون الأمر كذلك في كل عينة. أحياناً سوف يأخذ الباحث عينات من خلال فرصة عشوائية، قد لا تعكس تماماً المجتمعات التي سحبت منها هذه العينات. ومن هنا فإن النتيجة لمربع كاي ستأخذ قيمة موجبة:

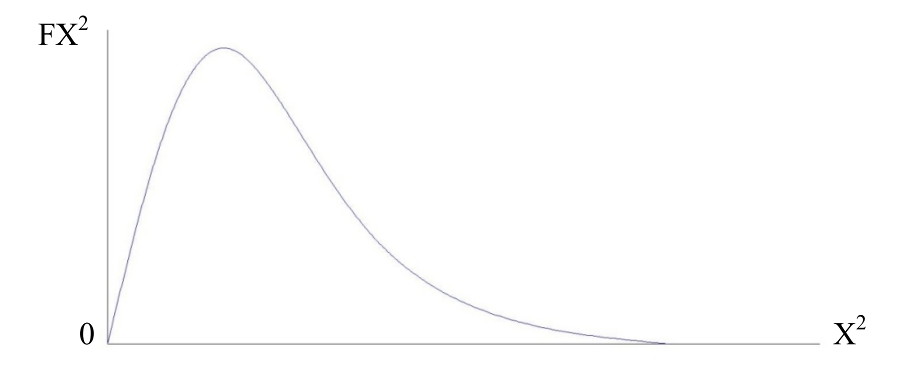
$$fo \neq fe \rightarrow (fo-fe)^2 > 0 \rightarrow X^2 > 0$$

وتجدر الإشارة إلى أنه كلما كان الفرق كبيراً بين التوزيعات التكرارية المشاهدة والتوزيعات التكرارية المتوقعة كانت قيمة مربع كاي كبيرة. لاحظ أنه في معادلة مربع

كاي، أن الفروق بين التوزيعات التكرارية المشاهدة، والتوزيعات التكرارية المتوقعة قد تم تربيعها. ويؤكد هذا التربيع أن المدى لكل القيم المحتملة لمربع كاي يجب أن تبدأ عند صفر وتزداد في الاتجاه الموجب، بغض النظر عما إذا كان التكرار المتوقع أكبر من التكرار المشاهد أو العكس بالعكس. إن تربيع أي فرق سوف يولد رقماً موجباً (بما أنه قد تم حساب مربع كاي على أساس الفرق بين الدرجات المتوقعة والدرجات الفعلية المربعة، وليس على أساس اتجاه الفرق حيث لا جدوى أن نختار بين الاختبار الاستدلالي أحادي الجانب أو ثنائي الجانب. فكل الفروق بين الدرجات المشاهدة والدرجات المتوقعة - بغض النظر عما إذا كانت هذه الفروق قد تكون ناشئة من الدرجات المشاهدة سواء كانت أعلى أو أدنى من الدرجات المتوقعة - سوف تأخذ قيمة موجبة)[2].

إن توزيع مربع كاي يحتوي على ذيل طويل انظر الشكل رقم (1)، الذي يعكس الحقيقة التي مفادها أن هناك احتمالية اختيار عينات عشوائية تعطي قيمة عالية لمربع كاي، بالرغم من أن المتغيرات تكون مستقلة، إلا أن هذا الأمر قد يكون بعيد الاحتمال. إن هذا الأمر سيكون مجرد ضربة حظ كي يختار الباحث عينة من مجموعة واحدة تكون الحالات فيها قد جاءت من نهاية طرف واحد من التوزيع، وإن عينة أخرى من مجموعة أخرى قد جاءت من نهاية الطرف الآخر من التوزيع إذا كان الفرض الصفري للاستقلال صحيحاً. وعليه، فإن المساحة تحت المنحنى للقيم الكبيرة لكاي تربيع تكون صغيرة بحيث تعكس احتمالية منخفضة لما حدث بالصدفة[3].

(لاحظ من خلال توزيع المعاينة لمربع كاي أنه يمكننا تحديد الاحتمالية بأن الفرق بين الدرجات المشاهدة والدرجات المتوقعة تكون ناتجة عن التباين العشوائي عند عملية المعاينة من مجتمعات يكون فيها المتغيران مستقلين). فعلى سبيل المثال، من الممكن أن نجد عينة (3) التي تناولناها سابقاً سوف تسحب مرة واحدة في الألف ($P = 0.001$) إذا كان المتغيران مستقلين عن بعضهما البعض. إن مثل هذا الأمر يمكن التفكير فيه، ولكنه على نحو بعيد الاحتمال كمبرر لنا من أن فرضيتنا حول الاستقلال يجب التخلي عنها - في الواقع ثمة تبعية أو اعتماد dependence بين الدخل ومكان الإقامة.



شكل (18 ـ 1) توزيع X^2

دعنا الآن نتناول المثال المتعلق بالدخل ومكان الإقامة لكي نوضح إجراءات اختبار مربع كاي للاستقلال بشكل عملي. فالمسح النظري الذي تناولناه سابقاً والذي يحتوي على 323 فرداً من المناطق الريفية و 397 فرداً من المناطق الحضرية، مع توزيع الإجابات يوضحه الجدول (18 – 5).

جدول (18 ـ 5) مكان الإقامة حسب الدخل
(التوزيعات المشاهدة)

	الدخل		
مكان الإقامة	منخفض	عال	المجموع
ريفي	205 (%55)	118 (%34)	323 (%45)
حضري	167 (%45)	230 (%66)	397 (%55)
المجموع	372 (%100)	348 (%100)	720 (%100)

نرى من خلاله أن أول رقم في كل خلية يشكل العدد الحقيقي لذوي الدخل المنخفض وذوي الدخل العالي الذين يعيشون في كل منطقة من هاتين المنطقتين، وتمثل 55 % كل الذين سجلوا بأن دخولهم منخفضة، ويعيشون في مناطق ريفية حيث يصل

عددهم إلى 205 فرداً. وعلى الجانب الآخر، فإن نسبة 34 % فقط من ذوي الدخول العالية يعيشون في مناطق ريفية (118 فرداً).

وتجدر الإشارة إلى أنه من خلال هذه البيانات يمكننا أن نرى الفرق بين أولئك الأفراد ذوي الدخول المنخفضة والأفراد ذوي الدخول العالية فيما يتعلق بالمناطق التي يعيشون فيها. إن السؤال الذي يمكن طرحه هنا هو: هل هذا الفرق يكون راجعاً إلى التباين العشوائي؟ للإجابة على هذا السؤال نحتاج، بادئ ذي بدء، إلى حساب التوزيعات المتوقعة: أي أن الأعداد المتوقع الحصول عليها في كل خلية من خلايا الجدول إذا كان المتغيران مستقلين (انظر الجدول 18 – 6).

جدول (18 ـ 6) مكان الإقامة حسب مستوى الدخل (التوزيعات المتوقعة)

مكان الإقامة	الدخل		
	منخفض	عال	المجموع
ريفي	$167.4 = 372 \times \frac{45}{100}$	$156.6 = 348 \times \frac{45}{100}$	% 45
حضري	$204.64 = 372 \times \frac{55}{100}$	$191.4 = 348 \times \frac{55}{100}$	% 55
المجموع	372	348	% 100

إن عدد المبحوثين المتوقع أن نجدهم في كل خلية إذا كان المتغيران مستقلين، تم حسابهم في كل خلية. فعلى سبيل المثال، إذا كان 45 % لكل المبحوثين يعيشون في مناطق ريفية، فإننا نتوقع أن نجد 45 % من ذوي الدخل المنخفض يعيشون في مناطق ريفية. يوجد لدينا عددٌ إجماليٌّ قدره 372 فرداً من ذوي الدخل المنخفض، 45 % من هؤلاء يعطينا 167 فرداً من ذوي الدخول المنخفضة المتوقع معيشتهم في مناطق ريفية وهكذا في باقي الخلايا (انظر جدول 18 – 7).

جدول (18 ـ 7) مكان الإقامة حسب مستوى الدخل (التوزيعات المشاهدة والتوزيعات المتوقعة)

مكان الإقامة	الدخل: منخفض	الدخل: عال	المجموع
ريفي	205 (167.4)	118 (156.6)	323
حضري	167 (204.6)	230 (191.4)	397
المجموع	372	348	720

يتبين من خلال الجدول أعلاه أن هناك فرقاً بين التوزيعات المشاهدة والتوزيعات التكرارية في كل خلية، وأنه بإمكاننا استخدام معادلة مربع كاي لبيان هذا الفرق في رقم واحد. ويوضح الجدول (18 – 8) إجراء حساب مربع كاي من هذه الفروق.

جدول (18 ـ 8) حساب قيمة مربع كاي (X^2)

مكان الإقامة	الدخل: منخفض	الدخل: عال	المجموع
ريفي	$X^2 = \frac{205-(167.4)^2}{167.4} = 8.5$	$X^2 = \frac{118-(156.6)^2}{156.6} = 9.5$	323
حضري	$X^2 = \frac{167-(204.6)^2}{204.6.4} = 6.9$	$X^2 = \frac{230-(191.4)^2}{191.4} = 7.4$	397
المجموع	372	348	720

وبعد إجراء عملية حساب قيمة مربع كاي (X^2) لكل خلية من خلايا الجدول، يمكننا بعدئذٍ إضافة هذه القيم بعضها لبعض لنتحصل على القيمة النهائية لمربع كاي والجدول التقاطع ككل التي تعبر عن الحقيقة بأن العينة الفعلية لا تتطابق تماماً مع الفرض الصفري للاستقلال:

$$\text{العينة} \quad x^2 = \sum \frac{(Fo - Fe)^2}{Fe}$$

$$= \frac{(205-167.4)^2}{167.4} + \frac{(118-156.6)^2}{156.6} + \frac{(167-204.6)^2}{204.6} + \frac{(230-191.4)^2}{191.4}$$

$$= 8.5 + 9.5 + 6.9 + 7.8 = 32.7$$

توزيعات مربع كاي The Distribution of X^2:

بعد أن تحصلنا على قيمة مربع كاي مساوية لـِ 32.7، فالسؤال الذي يمكن طرحه هنا هو ماذا تعني هذه القيمة المحسوبة؟ إن هذه القيمة المحسوبة، في حد ذاتها لا تخبرنا بالشيء الكثير، وبمعزلٍ عن الحقيقة فإن هذه القيمة ليست مساوية لصفر، وعليه فهي تشير إلى أن هناك بعض الاعتماد dependence بين هذين المتغيرين في بيانات هذه العينة. والسؤال المطروح هو ما إذا كانت هذه النتيجة تدعونا إلى رفض الفرض الصفري للاستقلال بالاعتماد على احتمالية القيمة المتحَصَّل عليها لهذه العينة وهي (X^2 = 32.7) من المجتمع حيث إن هذين المتغيرين مستقلان. ولتحديد هذه الاحتمالية يمكننا الإشارة إلى الجدول المتعلق بالقيم الحرجة لتوزيعات مربع كاي (انظر الملحق - جدول رقم 1 / 6). وباستخدام هذا الجدول يمكننا الحصول على القيمة الاحتمالية لمربع كاي آخذين في الاعتبار درجات الحرية. ولحساب درجة الحرية، ينبغي اتباع القاعدة التالية:

$$Df = (r-1)\ (C-1)$$

حيث أن : (r) تشير إلى عدد الصفوف.

: (c) تشير إلى عدد الأعمدة.

وفي جدول 2 × 2 كما هو الحال في المثال الذي بين أيدينا فإن درجة الحرية هي (1). وإن القيمة الحرجة المناظرة هي 3.841، على مستوى الدلالة 0.05:

$$x^2 = 3.841(\alpha = 0.05, df = 3)7.815$$

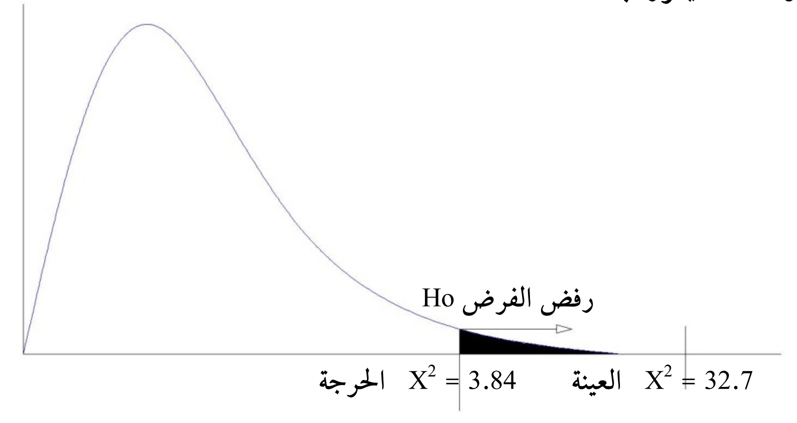
وتجدر الإشارة إلى أنه يمكن رفض الفرض الصفري H0، عندما تكون قيمة X^2 المحسوبة مساوية أو أكبر من قيمة X^2 المتوقعة. في المثال الذي يبين أيدينا، والذي اعتمدنا

قيمة الفا 0.05 بدرجة حرية واحدة، فإن القيمة الاحتمالية لـِ X^2 تساوي 3.841، في حين أن قيمة X^2 المحسوبة للعينة المدروسة هي 32.7. إذاً من خلال هذه المعلومات، يمكننا الوصول إلى نتيجة مفادها: أن القيمة الاحتمالية لمربع كاي تساوي 3.841، وأن القيمة المحسوبة (قيمة العينة) تساوي 32.7، وهي قيمة قد وقعت في المنطقة الحرجة (منطقة الرفض)، وعليه، يمكننا في هذا المثال، رفض الفرض الصفري H0 للاستقلال (مكان الإقامة والدخل ليسا مستقلين).

تجدر الإشارة إلى ملاحظة مهمة، وهي أن اختبار مربع كاي في حد ذاته لا يخبرنا عما تكون عليه طبيعة هذه العلاقة. وأن كل ما يمكن أن نقربه من خلال استخدام هذا الاختبار أن هناك بعضاً من الارتباط بين هذين المتغيرين، وقد اخترنا وصفاً لهذه العلاقة أنها علاقة ذات اتجاه واحد من الدخل إلى مكان الإقامة.

ربما أن شخصاً ما قد يدعي بسهولة أن السببية يمكن أن تتخذ اتجاهاً آخراً، وأن هناك علاقة متبادلةً معتمدةً بين هذين المتغيرين Mutual dependence. إن اختبار مربع كاي لا تقرر لنا مثل هذه القضية، بقدر ما يخبرنا بأن هذين المتغيرين ليسا مستقلين. بمعنى أنهما يرتبطا بعضهما ببعض[4].

في هذا السياق إذاً، كيف لنا أن نختار وصفاً لهذه العلاقة. إن الإجابة على ذلك تعتمد في الأساس على مسألة الجدال النظري الذي يقر بأن التحليل الإحصائي يستطيع أن يخبر ولكنه لا يقرر أبداً[5].



شكل (18 ـ 2) درجات القيم الحرجة والعينة

إجراء اختبار مربع كاي باستخدام برنامج SPSS:

توليد جداول التقاطع Crosstabs لمربع كاي باستخدام Spss:

1- من القائمة الموجودة في الجزء من الشاشة اختر:

Analyze ⇐ Summarize ⇐ Crosstabs

2- انقر على المتغير في القائمة التي تشكل صفوف الجدول في هذه الحالة Place of Residence.

3- انقر على ◄ الذي يشير إلى القائمة المحددة بعنوان Row(S)، تقود هذه العملية إلى لصق (Pastes) Place of Residence في القائمة المحددة لـِ Row(S) .

4- انقر على المتغير في القائمة التي تشكل أعمدة الجدول في هذه الحالة Income Level.

5- انقر على ◄ الذي يشير إلى القائمة المحددة بعنوان Column(s). هذه العملية تقود إلى لصق Income Level في القائمة المحددة لـِ Column(s).

6- انقر على زر Statistics، هذا يقود إلى إعطائنا مربع Statistics: Crosstabs.

7- اختر ch-Square بالنقر على الصندوق القريب منه بوضع علامة T في الصندوق للتأكد بأنه قد تم اختياره.

8- انقر على Continue.

9- انقر على OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

Crosstabs

Case Processing Summery

	CASES					
	Valid		Missing		TOTAL	
	N	Percent	N	Percent	N	Percent
Place of Residence Income Level	720	100.0	0	%0	720	%.0100

Place of Residence* Income Level Crosstabulation

		Income Level		TOTAL
		Low	High	
Place of Residence*	urban count	167	230	397
	within Income Level%	44.9%	66.1%	55.1%
	Rural Count	205	118	323
	within Income Level%	55.1%	33.9%	44.9
ToTAL	Count	372	348	720
	within Income Level%	100.0%	100.0%	100.0%

Chi - Square Tests

	Valne	Df	Asymp. sig (2-tailed)	Exact sig (2-tailed)	Exact sig (1-tailed)
Pearson Chi - Square	32.6676[b]	1	.000		
Contiuity Correction[a]	31.816	1	.000		
Likeli hood Ratio	32.965	1	.000		
Fisherś exact test				.000	.000
Linear - by Linear Association	32.622	1	.000		
N. of Valid Cases	720				

.a. Completed Only for a 2X2 table

.b. O Cells (0%) have Expected Count is 156.12

المصدر: George Argyrous, Statistics for Social & Health Research, op.cit , p.406.

شكل (18 ـ 3) مخرجات SPSS لمربع كاي

تفسير مخرجات SPSS لمربع كاي (X^2):

إن الشيء المهم الذي ينبغي على الباحث الاهتمام به من هذه المخرجات هي قيمة Pearson chi-Square الموجودة في المربع الأخير المُعَنْوَن باسم chi-Square tests (اختبار مربع كاي). في الصف الأول المُعَنْوَن بـ Pearson chi-Square تحت قيمة 32.667 التي تشير إلى قيمة X^2 للعينة التي تم حسابها سابقاً يدوياً، بدرجة حرية 1 عمود (3)، ودلالة 000. (عمود 4). وبالرغم من أن درجة الدلالة هذه تعني أقل من ألفا 0.05 في 10.000 للحصول على عينات لهذا التوزيع من المجتمعات التي تكون فيها المتغيرات مستقلة. ومن خلال هذه القيمة الاحتمالية المنخفضة فهي تقودنا إلى رفض الفرض الصفري للاستقلال.

أما إذا كان لديك جدول 2 × 2، وكذلك عندما تكون أي خلية من خلايا التوزيعات النظرية صغيرة جداً، أي أقل من 10 حالات حينئذ على الباحث أن يستخدم القيمة الموجودة في الصف الثاني من مربع chi-Square tests، وهذه القيمة هي قيمة Contiuity Correction التي وضعها yates (تعد هذه القيمة تعويضاً للمبالغة في تقدير اختبار X^2 عند استخدام جدول 2 × 2).

وبالنظر إلى القيمة في عمود (Asymp. Sig. (2.tailed يمكن للباحث أن يتخذ القرار المناسب، فإذا كانت هذه القيمة الموجودة في هذا العمود أكبر من قيمة 0.05، عندئذٍ يمكن للباحث التوصل إلى أن النتيجة ليست دالة والعكس بالعكس.

مثال لزيادة التوضيح:

يمكننا الآن العمل من خلال هذا المثال لزيادة التوضيح مستخدمين الخطوات الخمس المتبعة في إجراء اختبار الفروض. ومن خلال البيانات المبينة في الجدول التالي لعدد 800 طفلاً، فيما يتعلق بعلاقة النوع، وما إذا كان هؤلاء الأطفال يشاهدون الأخبار في الإذاعة المرئية.

جدول (18 ـ 9) مشاهدة الأطفال للأخبار في الإذاعة المرئية

مشاهدة الأخبار في الإذاعة المرئية	النوع		
	إناث	ذكور	المجموع
نعم	377	363	740 (92 %)
لا	25	35	60 (8 %)
المجموع	402	398	800

المصدر: George Argyrous, Statistics for Social & Health Research, op.cit , p.414.

الخطوة الأولى: صياغة الفرض الصفري والفرض البديل.

H0: النوع ومشاهدة الأخبار المرئية متغيران مستقلان عن بعضهما البعض.

Hi: النوع ومشاهدة الأخبار المرئية متغيران غير مستقلين عن بعضهما البعض.

الخطوة الثانية: اختيار اختبار الدلالة.

سنقوم في هذا المثال بتحليل بيانات عينة تم ترتيبها في جدول ثنائي لمعرفة ما إذا كانت هناك علاقة بين هذين المتغيرين اللذَيْن نجري اختبار X^2 للاستقلال عليهما كاختبار ملائم للاستدلال.

الخطوة الثالثة: حساب درجة العينة:

ولحساب درجة العينة يتعين علينا بادئ ذي بدء حساب التوزيعات المتوقعة استناداً إلى مجموع نسب الأعمدة في الجدول (9).

جدول (18 ـ 10) مشاهدة الأطفال للأخبار المرئية وفقا للنوع

مشاهدة الأخبار في الإذاعة المرئية	النوع		
	إناث	ذكور	المجموع
نعم	371.8	368.2	740
لا	30.2	29.8	60
المجموع	402	398	800

$$x^2 = \sum\frac{(Fo - Fe)^2}{Fe}$$

$$= \frac{(377-371.8)^2}{371.8} + \frac{(363-368.2)^2}{368.2} + \frac{(25-30.2)^2}{30.2} + \frac{(35-29.8)^2}{29.8}$$
$$= 1.9$$

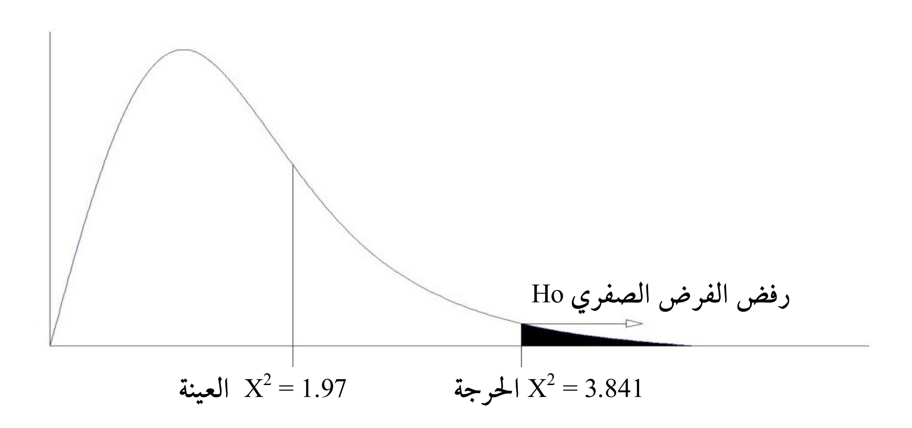
الخطوة الرابعة: إيجاد الدرجة الحرجة والمنطقة الحرجة:

في هذا المثال، نحن نتعامل مع جدول 2 × 2، وعليه، فإننا نتحصل على درجة حرية واحدة على مستوى ألفا 0.05 = α، وبالتالي فإن القيمة الحرجة لِـ X^2 تساوي =

$$x^2 = 3.841(\alpha = 0.05)df = 1$$

الخطوة الخامسة: اتخاذ القرار:

من خلال الشكل (18 – 4) فإن درجة العينة لا تقع في المنطقة الحرجة، وعليه، لا يمكننا رفض الفرض الصفري بأن هذين المتغيرين مستقلان.



شكل (18 ـ 4) الدرجة الحرجة ودرجة العينة

الافتراضات والقيود في استخدام مربع كاي:

لكي يُسْتَخْدَمَ مربع كاي لجودة التطابق أو اختيار مربع كاي للاستقلال، توجد مجموعة من الشروط ينبغي توفرها. وأن أي اختيار إحصائي يخالف هذه الافتراضات والقيود يبعث بالشك في النتائج التي يتوصل إليها الباحث.

على سبيل المثال، إن الاحتمالية للوقوع في الخطأ من النوع الأول Type I Error يمكن أن يُشَوَّه عندما لا يفي بافتراضات الاختبارات الإحصائية.

هناك بعض الافتراضات والقيود ينبغي مراعاتها عند استخدام اختبارات مربع كاي:

1- يستند مربع كاي على فرضية أن تكون العينات خاضعة للشروط الاحتمالية. ومن هنا يكون من غير المناسب استخدام مربع كاي أو أي مقاييس استدلال لعينات لم يتم حسابها باستخدام العينات الاحتمالية.

2- ينبغي أن تكون المشاهدات مستقلة عن بعضها البعض. في هذه النقطة ينبغي ألا يخلط بينها وبين مفهوم الاستقلالية بين المتغيرات كما رأينا عند الحديث عن اختبار الاستقلال. أن أحد نتائج المشاهدات المستقلة تكون بأن كل تكرار مشاهد يتولد من

خلال موضوع مختلف. بمعنى آخر، أن المشاهدات المستقلة يمكن حسابها من كل فرد أو حالة مرة واحدة فقط، حيث لا يمكن أن يظهر الفرد أو الحالة في أكثر من فئة أو مجموعة. كما أن البيانات المأخوذة من فرد واحد لا يمكن أن تؤثر في المشاهدات المأخوذة من حالة أخرى أو فرد آخر. ومن هنا يكون اختبار مربع كاي اختباراً غير ملائم عندما يكون شخص ما قد أنتج إجابات يمكن تصنيفها في أكثر من فئة أو تسهم في أكثر من حساب توزيع لفئة مفردة بمعنى، أن كل ملاحظة أو قياس لا يكون متأثراً بالملاحظة الأخرى. فالملاحظة أو القياس الذي نتحصل عليه من كل فرد ينبغي إلا يؤثر على البيانات المأخوذة من عينة أو حالة أخرى.

3- حجم التكرارات المتوقعة: تجدر الإشارة إلى أنه ينبغي على الباحث ألا يستخدم اختبار مربع كاي عندما يكون التوزيع التكراري لأي خلية في التوزيع أقل من 5. لأن إحصاء مربع كاي يمكن أن يُشَوَّه عندما يكون التكرار المتوقع صغيراً جداً. انظر على سبيل المثال، لو أنك قمت بحسابات مربع كاي لخلية مفردة. نفرض أن الخلية بها قيمة fe=1 وأن fo=5. فإن حساب هذه الخلية للقيمة الكلية لمربع كاي تكون كالتالي:

$$\text{(الخلية) Cell} = \frac{(f0 - fe)^2}{fe} = \left(\frac{5-1}{1}\right)^2 = \frac{4^2}{1} = 16$$

والآن يمكنك التفكير في مثال آخر تكون فيه Fe= 10 و F0= 14 إن الفرق بين التوزيعات التكرارية المشاهدة، والتوزيعات التكرارية المتوقعة لازال (4)، إلا أن إسهام هذه الخلية للقيمة الكلية لمربع كاي تختلف عن الحالة الأولى:

$$\text{(الخلية) Cell} = \frac{(f0 - fe)^2}{fe} = \frac{(14-10)^2}{10} = \frac{4^2}{10} = 1.6$$

من هنا يتضح جلياً أن قيمة fe صغيرة لديها تأثير كبير على قيمة مربع كاي. وتصبح هذه المشكلة أكثر حدة عندما تكون قيمة fe أقل من 5. وعندما تكون fe صغيرة جداً ستولد قيمة عالية. ومن هنا نجد أن اختبار مربع كاي اختبار حساس عندما تكون قيم fe صغيرة بشكل مفرط. ولتحاشي هذه المواقف، ينبغي على الباحث استخدام عينات كبيرة بدلاً من العينات الصغيرة[6].

قد يصادف الباحث عند تحليل البيانات خلايا مختلفة لديها توزيعات تكرارية أقل من 5. وكنتيجة لذلك، فإن قيمة مربع كاي يمكن أن يكون مغالياً فيها، مما يؤدي إلى رفض الفرض الصفري بالرغم من عدم وجود تطابق بين المتغيرين تحت الدراسة في المجتمع.

إن السؤال المطروح في هذا السياق هو: ماذا ينبغي على الباحث فعله كي يتغلب على هذه المشكلة؟

للتغلب على مثل هذه المشكلة يكون بإمكان الباحث - في بعض الأحيان - اختزال بعض الفئات لواحد من المتغيرين أو كليهما في جدول التوافق لتوليد تكرارات مشاهدة وتكرارات متوقعة كبيرة داخل الخلايا. ومن ناحية ثانية، باستطاعة الباحث أن يجمع فئات المتغيرات، فقط عندما تكون هذه الفئات مرتبطة مع بعضها البعض ارتباطاً منطقياً. وبشكل عام، يمكن للباحث أن يجمع فقط تلك الفئات المتجاورة والمقاسة على متغيرات ترتيبية. فعلى سبيل المثال، يمكن للباحث أن يتعامل مع متغير "زيارة الأقارب"، وذلك بجمع المبحوثين الذين يزورون أقاربهم مرة في العام، مع أولئك الذين أجابوا مرتين أو ثلاث مرات في العام. ولكن ليس من المعقول أن يجمع الباحث إجابات المبحوثين الذين يزورون أقاربهم مرة في الشهر أو مرة في الأسبوع لمجرد أن يحصل الباحث على تكرارات مشاهدة وتكرارات متوقعة كبيرة. كذلك الأمر إذا كان لدى الباحث مجموعات منفصلة مثل: منخفض، منخفض جداً، عال. ففي هذه الحالة يمكن للباحث أن يختزل هذا المقياس إلى فئتين: (منخفض + منخفض جداً) وعال ليصبح المقياس: منخفض وعال.

أما فيما يتعلق بجداول التوافق الصغيرة التي تحتوي على فئتين فقط للمتغير المستقل، والمتغير التابع (جداول 2×2) بدرجة حرية 1 فقط، فإنه ليس بالإمكان استخدام مربع كاي إذا كانت أي من التوزيعات المتوقعة للخلية تحتوي دون 5. أما إذا كانت توزيعات خلية 5 أو أكثر، ولكنها أقل من 10. ففي هذه الحالة يتوجب على الباحث إجراء بعض التعديلات الطفيفة على معادلة مربع كاي، وذلك باستخدام تصحيح ياتس yateś Correction:

$$x^2 = \sum \frac{(|fo - fe| - 0.5)^2}{fe}$$

لاحظ هنا، أننا نطرح 5، من الفرق بين التكرارات المشاهدة والتكرارات المتوقعة لتوزيعات الخلية قبل تربيع النتيجة. ويهدف هذا التعديل إلى التقليل من التأثير على قيمة مربع كاي في الفرق بين التكرارات المشاهدة والتكرارات المتوقعة في كل خلية على حدة.

ثالثاً: من المشكلات الأخرى المرتبطة باستخدام مربع كاي هي تأثر مربع كاي بحجم العينة الكبيرة (مشكلات العينة الكبيرة). فعندما يقوم الباحث بحساب مربع كاي في جدول التوافق مبيناً وجود تطابق ضعيف بين متغيرين في عينة صغيرة، ففي هذه الحالة، فإن قيمة مربع كاي، على الأرجح أنها تشير إلى تطابق يمكن مشاهدته. وأن هذه القيمة المتحصل عليها قد جاءت نتيجة للصدفة. وفي حالة أخرى، عندما يحصل الباحث على تطابق بنفس القوة في عينة كبيرة، فإن قيمة مربع كاي تصبح أكبر. وكنتيجة لذلك، فإن التطابق الذي تم الحصول عليه في العينة الصغيرة، ربما سيؤدي إلى احتمالية قبول الفرض الصفري. أما التطابق الذي تم الحصول عليه في العينة الكبيرة سوف يؤدي إلى احتمالية رفض الفرض الصفري، بالرغم من أن قوة التطابق هي واحدة.

تجدر الإشارة إلى أنه ثمة مَنْ يعتقد من الباحثين أن هذه المشكلة في استخدام مربع كاي تؤدي بهم إلى رفض الفرض الصفري بوجود هذه التطابقات الطفيفة. وكنتيجة لذلك، نلاحظ في كثير من التقارير العلمية أنها تحتوي على تطابق ذي دلالة إحصائية، بالرغم من أن هذا التطابق ضعيف جداً، فإنه من الأرجح أن تكون هذه النتيجة قد جاءت بالصدفة.

في حين يعتقد فريق آخر من الباحثين أن حجم العينة لا تشكل مشكلة في استخدام مربع كاي على الإطلاق. ومن هنا يرى هذا الفريق أنه كلما زاد حجم العينة زادت معه ثقتنا في أن التطابق الذي سنتحصل عليه من هذه العينات سيكون ممثلاً للمجتمع الذي سحبت منه هذه العينات.

إن هؤلاء الباحثين قد ميزوا بين القضايا المتعلقة بالدلالة العملية Practical Significance عن القضايا المتعلقة بالدلالة الإحصائية Statistical Significance. فقد تكون درجة التطابق ذات دلالة إحصائية، لكن ذلك لا يعني البتة أنها ذات دلالة عملية.

ففي العينات الكبيرة حتى درجات التطابق الصغيرة للغاية قد تصل إلى درجة الدلالة الإحصائية. فإذا كانت درجة التطابق قيمتها صغيرة ولها دلالة إحصائية، لكن دلالتها العملية محدودة جداً(7).

اختبار أحادي الجانب واختبار ثنائي الجانب :

إن اختبار أحادي الجانب واختبار ثنائي الجانب الذي تناولناه في متن فصول هذا الكتاب، لا تهمنا كثيراً عند التعامل مع اختبار مربع كاي للاستقلال، وعندما يكون الفرض الذي نود اختباره اتجاهياً، فعلى الباحث استخدام اختبار ثنائي الجانب عند التعامل مع هذا الاختبار.

قياس حجم التأثير لاختبار مربع كاي للاستقلال :

إن اختبار أي فرض مثله مثل اختبار مربع كاي للاستقلال يسعى لتنظيم الدلالة الإحصائية للنتائج التي يتوصل إليها من خلال دراسة ما. وبالتحديد فإن الهدف من وراء أي اختبار هو تحديد ما إذا كانت أنماط أو علاقات المشاهدة في بيانات العينة يرجح أن تكون قد حدثت بالصدفة؛ وهذا يعني بدون أي أنماط متشابهة أو علاقات في المجتمع. أن اختبارات الدلالة عادة لا تتأثر فقط بالحجم، أو قوة تأثيرات المعالجة، ولكنها تتأثر أيضاً بحجم العينات. وكنتيجة لذلك، فإن وجود أي تأثير بسيط يمكن أن يكون ذا دلالة إحصائية إذا ما تَمَّتْ ملاحظتُهُ من خلال عينة كبيرة جداً. وحيث إن تأثير الدلالة ليس بالضرورة أن يعني تأثيراً كبيراً، فإنه بشكل عام. يُنْصَحُ بأَنَّ نتائج اختبار فرض ما يجب أن يصاحبه قياس حجم التأثير. إن هذه النصيحة العامة، يجب أيضاً أن تطبق على اختبار مربع كاي للاستقلال.

معامل فاي ϕ و Cramerś V :

يعتبر معامل فاي ϕ مقياساً للارتباط للجدول الثنائي للبيانات التي تحتوي على متغيرين ثنائيين (كلا المتغيرين يأخذان قيمتين فقط مثل: النوع، ذكر / أنثى... الخ)،

وتنطبق هذه الحالة أيضاً عند إجراء اختبار مربع كاي لمصفوفة تحتوي على 2×2 (مرة أخرى كلا المتغيرين يأخذان قيمتين فقط). في مثل هذه الأحوال يمكننا حساب معامل ϕ، بالإضافة إلى حساب اختبار مربع كاي للاستقلال لنفس البيانات. ولما كانت معامل فاي ϕ مقياساً لقوة العلاقة بدلاً من مقياس الدلالة (تستخدم المعادلة التالية لاختبار دلالة معامل فاي ϕ: $(X^2 = N(\phi^2)$)، وبالتالي فهي تقدم لنا مقياس حجم التأثير.

وتجدر الإشارة إلى أن قيمة معامل فاي يمكن حسابها مباشرة من مربع كاي وفقاً للمعادلة التالية:

$$\phi = \sqrt{\frac{x^2}{N}}$$

وتُحَدَّد قيمة معامل فاي بشكل كلي من خلال النسب في مصفوفة البيانات 2×2؛ وهي بالتالي مستقلة تماماً عن الحجم المطلق للتكرارات.

إن قيمة مربع كاي، على أي حال، تتأثر بالنسب وكذلك حجم التكرارات.. إن هذا الفرق يمكن البرهنة عليه من خلال المثال التالي:

مثال:

البيانات التالية تبين توزيع تكراري يقيم العلاقة بين النوع والاتجاه نحو تنظيم الأسرة:

الاتجاه	النوع	
	ذكر	أنثى
أوافق	5	10
لا أوافق	10	5

لاحظ أن هذه البيانات تظهر أن الإناث تفضل تنظيم الأسرة أكثر منه لدى الذكور. ومن خلال هذه البيانات يمكننا الحصول على قيمة مربع كاي مساوية لـِ 3.33 (وهي قيمة غير دالة)، وكذلك قيمة معامل فاي $\phi = 0.3332$. وفي الخطوة التالية نحتفظ بنفس

النسبة تماماً لهذه البيانات ولكننا سنضاعف كل التكرارات. وتظهر هذه العملية في المصفوفة التالية:

الاتجاه	النوع	
	ذكر	أنثى
أوافق	10	20
لا أوافق	20	10

مرة أخرى، نجد أن الإناث يفضلن تنظيم الأسرة، إذا ما قورن ذلك بالذكور. الآن، يمكننا ملاحظة أن العينة تحتوي على 30 من الإناث و30 من الذكور. ويمكننا الحصول على قيمة مربع كاي وفاي من هذه البيانات الجديدة: قيمة مربع كاي تساوي 6.66 وهي قيمة ضعف قيمة مربع كاي في المرة الأولى (وهي الآن قيمة ذات دلالة على مستوى ألفا $\alpha = 0.05$). ولكن قيمة فاي بقيت كما هي دون تغير 0.3332.

ولما كانت النسبة متساوية في كلتا العينتين، فإن قيمة معامل فاي لم تتغير. وعليه يمكننا القول بأنه كلما كانت العينة كبيرة، تقدم أكثر حدث مقنع إذا ما قورنت بالعينة الصغيرة، وأكثر ترجيحاً لتمدنا بنتيجة دالة.

ولتفسير قيمة فاي يمكن للباحث أن يتبع نفس المعيار المستخدم في تقييم الارتباط. وتعتبر قيمة $\phi = 0.10$ (تأثير متوسط) و 0.30 تأثير متوسط و 0.50 تأثير كبير. في بعض الأحيان تربع قيمة ϕ (ϕ^2) لتبين لنا نسبة التباين الذي يمكن وضعه في الاعتبار، كما هو الحال عند التعامل مع قيمة r^2. أما إذا كان اختبار مربع كاي يتعلق بمصفوفة تحتوي على أكثر من 2×2، فإن تعديلاً لمعامل فاي ينبغي إجراؤه. ويُعرف هذا التعديل بـِ: Cramerś V لقياس حجم التأثير:

$$V = \sqrt{\frac{X^2}{n(df_*)}}$$

لاحظ معادلة Cramerś V مطابقة لمعادلة معامل فاي، باستثناء الإضافة المتعلقة بـِ (df*) في مقام المعادلة.

إن قيمة df* ليست مساوية لدرجة الحرية df لاختبار مربع كاي للاستقلال حيث إن df = (C-1) (R-1). وتشير R إلى عدد الصفوف، و C إلى عدد الأعمدة. أما فيما يتعلق بـِ: Cramerś V، فإن قيمة df* تشير إلى أيهما أصغر (R-1) أو (C-1). العَلاَّمة كوهين Cohen في عام 1988 قد اقترح معايير لتفسير معامل Cramerś V كما يوضحه الجدول التالي:

جدول (18 ـ 11): معايير تفسير معامل Cramerś V

درجة الحرية	حجم التأثير		
	تأثير بسيط	تأثير متوسط	تأثير كبير
df*=1	0.10	0.30	0.50
df*=2	0.07	0.21	0.35
df*=3	0.06	0.17	0.29

المصدر: J. Cohen , Statistical Power analysis for behavioral Sciences, SAN Diego, CA: Academic Press , 1988.

لاحظ عندما تكون df*=1 كما هو الحال في مصفوفة 2×2، فإن معيار تفسير V بالضبط، هو مساوٍ لمعيار تفسير الارتباط المألوف أو معامل فاي. وللتدليل على ذلك، يمكننا الاستعانة بالبيانات التالية التي تقيم العلاقة بين النوع والرغبة في الاستفادة من خدمات الصحة النفسية.

جدول (18 ـ 12): الرغبة في الاستفادة من خدمات الصحة النفسية

النوع	الاحتمالية نعم	ممكن	الاحتمالية لا	مج
ذكور	11	32	17	60
إناث	34	43	13	90
مج	45	75	30	N=150
$x^2 = 8.23\ df = 2\ (5.9□)\ \propto = .05$				

المصدر: Frederick J. Gravetter and Larry B. Wallnau , Statistics for the behavioral Sciences , . op..cit , P. 623

لدينا مصفوفة 2×3 بعدد إجمالي 150 = N مشارك. وقد أنتجت هذه البيانات مربع كاي يساوي 238. = x^2، وباستخدامنا هذه القيمة نتحصل على:

$$V = \sqrt{\frac{x^2}{ndf_*}} = \sqrt{\frac{8.23}{150(1)}} = \sqrt{0.055}$$

$$= 0.23$$

وطبقاً لمعايير كوهين، فإن هذه القيمة المتحصل عليها تشير إلى علاقة صغيرة[8].

القوة Power:

يمكن تحديد القوة من الجدول التالي مستخدمين معلومات حجم التأثير المتمثلة في الحجم الكلي لدرجة الحرية TOTAL df وحجم العينة (N) (TOTAL 'N')، إن النتيجة التي تحصلنا عليها في المثال السابق أشارت إلى تأثير متوسط الحجم بدرجة حرية مساوية (1). (df = 1). وبالرغم من أن الجدول يحتوي على 200 = N فقط. وحتى هذه القيمة لِـ (N)، فإن القوة تصل إلى 97. بحجم كلي لدرجة الحرية يساوي (2) (للتعامل مع جدول القوة المتوقعة يستخدم الحجم الإجمالي لدرجة الحرية).

من الواضح أن القوة في هذا المثال حقاً عالية جداً، وبالطبع فهي قوة مقبولة. وتعني القوة في هذا المثال، أننا مفترضون أن الفرض الصفري فرضُ صحيحُ. إن فرصة 3% تقع

إذا ما أعيد بنفس الحجم (N). ولما كانت حسابات حجم التأثير والقوة حسابات علمية دقيقة فإن الباحث يمكنه الرجوع إلى جداول كوهين التفصيلية[9].

جدول (18-13): قوة اختبار مربع كاي X^2 على مستوى دلالة $\alpha = 0.05$ و α 0.01 بدرجة حرية 1، 2، 3، 4، 5

درجة الحرية df	الحجم الكلي للعينة (N)	حجم التأثير $\alpha = 0.05$			حجم التأثير $\alpha = 0.01$		
		كبير	متوسط	صغير	كبير	متوسط	صغير
1	25	.70	.32	.08	.47	.14	.02
	50	.94	.56	.11	.83	.32	.03
	100	*	.85	.17	.99	.66	.06
	200	*	.99	.29	*	.95	.12
2	25	.60	.25	.07	.36	.10	.02
	50	.90	.46	.09	.74	.24	.02
	100	*	.77	.13	.98	.55	.04
	200	*	.97	.23	*	.91	.08
3	25	.54	.21	.07	.30	.08	.01
	50	.86	.40	.08	.68	.19	.02
	100	.99	.71	.12	.97	.48	.03
	200	*	.96	.19	*	.87	.07
4	25	.50	.19	.06	.26	.07	.01
	50	.82	.36	.08	.62	.16	.02
	100	.99	.66	.11	.96	.43	.03
	200	*	.94	.17	*	.84	.06
5	25	.45	.17	.06	.23	.06	.01
	50	.79	.33	.07	.58	.14	.02
	100	.98	.62	.10	.94	.38	.03
	200	*	.93	.16	*	.80	.05

* الدرجة قريبة جداً من 1 صحيح.

المصدر: J. Cohen , Statistical Power analysis for the behavioral Sciences ,SAN Diego , CA: Academic Press , 1988 , PP. 228 - 230 and PP. 235 - 237.

تطبيقات خاصة لاختبارات مربع كاي : [10]

في بداية هذين الفصلين قد قدمنا اختبارات مربع كاي كأمثلة للاختبارات غير البارامترية. وبالرغم من أن الاختبارات غير البارامترية اختبارات تؤدي وظيفة فريدة خاصة بهذه الاختبارات، إلا أنه يمكن النظر إلى هذه الاختبارات كبديل لتقنيات الاختبارات البارامترية (المعلمية) الشائعة التي تم مناقشتها في متن هذا الكتاب.

وبشكل عام، تستخدم الاختبارات غير البارامترية كبديل للتقنيات المعلمية في مواقف قد يحدث في واحد من هذين الموقفين:

1- إن البيانات المجمعة قد لا تلبي الافتراضات المطلوبة لخصائص الاختبار المعلمي.

2- قد تحتوي البيانات على مقياس اسمي أو مقياس ترتيبي قد يصعب معه حساب معلمات المجتمع: كالمتوسط الحسابي، والانحراف المعياري.

في هذا الجزء، سنتناول بعض العلاقات بين اختبارات مربع كاي، والإجراءات البارامترية التي يمكن أن يُسْتَخْدَمَ فيها اختبار مربع كاي كبديل لهذه الإجراءات.

أ- مربع كاي وارتباط بيرسون (r).

ب- مربع كاي والمقاييس المستقلة لـِ وأنوفا ANOVA.

ج- معامل فاي ϕ.

د- اختبار الوسيط للعينات المستقلة.

أ- مربع كاي وارتباط بيرسون (r):

إن اختبار مربع كاي للاستقلال ومعامل ارتباط بيرسون (r) كلاهما تقنيات إحصائية تهدفان إلى تقييم العلاقة بين متغيرين. ويحدد نوع البيانات المتحصل عليها في دراسة معينة أي من الإجراءات الإحصائية لهذين النمطين يمكن أن يكونا ملائماً.

نفترض على سبيل المثال، أن باحثاً يرغب في استقصاء العلاقة بين تقدير الذات Self-Esteem والتحصيل الأكاديمي Academic Performance لعدد من الأطفال يبلغون من العمر عشر سنوات. فإذا تحصل الباحث على درجات عددية Numerical Scores لكلا المتغيرين، فإن نتائج البيانات ستكون مشابهة للقيم التالية:

المشاركون في الدراسة	تقدير الذات (X)	الأداء الأكاديمي
1	13	73
2	19	88
3	10	71
4	22	96
5	20	90
6	15	82
:	:	:
:	:	:

المصدر: Frederick J. Gravetter and Larry B. Wallnau , op. Cit , P. 630

وأن الباحث يمكنه أن يستخدم معامل ارتباط بيرسون (r) من أجل تقييم هذه العلاقة. وعلى الجانب الآخر، يمكن للباحث أن يختار، ببساطة، أن يصنف الأفراد إلى فئات لكلا المتغيرين. فعلى سبيل المثال، يمكن له أن يصنف كل طالب، إما إلى تحصيل عال أو تحصيل منخفض. وتصنيف متغير تقدير الذات إلى: تقدير عال، متوسط، ومنخفض.

إن نتيجة البيانات ستولد توزيعاً تكرارياً يمكن ملاحظته في مصفوفة مثل تلك المصفوفة الواردة في الجدول التالي:

جدول (18 ـ 14): توزيع تكراري لمستوى تقدير الذات وفقا للأداء الأكاديمي لعينة من 150 طفلا يبلغون من العمر عشر سنوات

	مستوى تقدير الذات			
الأداء الأكاديمي	عال	متوسط	منخفض	مج (N)
عال	17	32	11	60
منخفض	13	43	34	90
مج (N)	30	75	45	N=150

المصدر: .Ibid , P. 631

لاحظ أن هذه البيانات لا تتضمن أي درجات عددية. ولكنها تحتوي مجموعة تكرارات تتلاءم واختبار مربع كاي.

بـ مربع كاي والمقاييس المستقلة t وأنوفا:

مرة ثانية، يمكننا النظر إلى باحث يود استقصاء العلاقة بين تقدير الذات، والأداء الأكاديمي لمجموعتين من الأطفال يبلغون من العمر عشر سنوات. نفترض في هذه المرة، أن الباحث قام بقياس الأداء الأكاديمي بتصنيف الأفراد إلى فئتين (عال ومنخفض)، وبعد ذلك تحصل عل درجة عددية لتقدير الذات لكل طفل. وأن نتيجة البيانات ستكون مشابهة للدرجات الواردة في جدول (أ). وأن اختبار t للعينات المستقلة سيستخدم لتقييم فرق المتوسط بين المجموعتين من الدرجات. وكخيار بديل، يمكن للباحث أن يقيس تقدير الذات بتصنيف الأفراد إلى ثلاث فئات: عال، متوسط، منخفض. وإذا تحصل على درجة عددية للأداء الأكاديمي لكل فرد، فإن نتيجة البيانات تظهر في شكل الدرجات في الجدول (ب)، وأن اختبار أنوفا يمكن إجراؤه لتقييم الفروق في المتوسطات بين المجموعات الثلاثة.

جدول (18 ـ 15)

البيانات المناسبة للمقاييس المستقلة لاختبار t أو لاختبار أنوفا الجزء (أ) درجات تقدير الذات تم الحصول عليها من مجموعتين من الطلاب مختلفين في مستوى الأداء الأكاديمي. والجزء (ب) درجات الأداء الأكاديمي التي تم الحصول عليها لعدد ثلاث مجموعات مختلفين في مستوى تقدير الذات

(أ) :

درجات تقدير الذات لمجموعتين من الطلاب الأداء الأكاديمي	
عال	منخفض
17	13
21	15
16	14
24	20
18	17
15	14
19	12
20	19
18	16

(ب):

درجات الأداء الأكاديمي لثلاث مجموعات تقدير الذات		
عال	متوسط	منخفض
94	83	80
90	76	72
85	70	81
84	81	71
89	78	77
96	88	70
91	83	78
85	80	72
88	82	75

المصدر: .Ibid , P. 631

إن الهدف من وراء بيان هذه الأمثلة، هو أن اختبار مربع كاي للاستقلال، ومعامل ارتباط بيرسون (r)، واختبارات فروق المتوسط جميعها تقنيات إحصائية يمكن توظيفها لتقييم العلاقة بين متغيرين. إن التمييز الأساسي بين مختلف هذه الإجراءات الإحصائية يكمن في شكل البيانات. ومع ذلك، هناك تميزُ آخر هو الهدف الأساسي لهذه الإحصاءات المختلفة.

إن اختبار مربع كاي، واختبارات فروق المتوسط (t، وأنوفا) تستخدمان لتقييم دلالة

العلاقة، بمعنى أن هذه التقنيات تقرر ما إذا كانت العلاقة المشاهدة في العينة تمدنا بدليل كافي لنقرر أن هناك علاقة تشابه في المجتمع. وكذلك يمكن للباحث أن يقيم دلالة ارتباط بيرسون (r)، ومع ذلك، فإن الهدف الأساسي للارتباط هو قياس قوة Strength العلاقة. بالخصوص، إن تربيع قيمة r^2، تمدنا بقياس حجم التأثير Effect size وذلك بوصف نسبة التباين Proportion of Variance في أحد المتغيرين وما يحدثه في علاقته مع المتغير الآخر.

ج- معامل فاي ϕ :

لقد بينا في الفصول السابقة أن العلاقة بين متغيرين ثنائيين يمكن تقييمها من خلال، إما معامل فاي ϕ أو اختبار مربع كاي للاستقلال لمصفوفة 2×2. ففي المثال الذي تناولناه في الصفحات السابقة، واختبرنا فيه العلاقة بين النوع (ذكور/ إناث) وتفضيلها فريق أ أو ب. نلاحظ أن معامل فاي ϕ أنتجت ارتباطاً من شأنه أن يقيس قوة العلاقة. في حين أن اختبار مربع كاي يقيم دلالة هذه العلاقة.

د- اختبار الوسيط للعينات المستقلة:

يمدنا اختبار الوسيط بالإحصاءات غير البارامترية بديلة للمقاييس المستقلة لـِ t أو أنوفا ANOVA لتحديد ما إذا كان هناك فروق دالة بين اثنين أو أكثر من العينات المستقلة. ويصاغ الفرض الصفري لاختبار الوسيط على أن عينات عدة جاءت من مجتمعات تتقاسم وسيطاً مشتركاً (لا توجد فروق). في حين يصاغ الفرض البديل على أن العينات جاءت من مجتمعات عدة لا تتقاسم وسيطاً مشتركاً.

إن المنطق وراء اختبار الوسيط هو أنه متى يتم اختيار عدة عينات مختلفة من نفس توزيع المجتمع، فإن نصف الدرجات تقريباً في كل عينة ينبغي أن تكون هذه الدرجات فوق وسيط المجتمع، والنصف الآخر من الدرجات ينبغي أن يكون تحت الوسيط. بمعنى أن كل العينات المنفصلة ينبغي أن تكون موزعة حول نفس الوسيط. وفي الجانب الآخر، إذا كانت العينات قد جاءت من مجتمعات لديها وسيطات مختلفة، حينئذٍ فإن الدرجات في بعض العينات سوف تكون وبشكل متساوق أعلى من الدرجات في عينات أخرى، وبالتالي ستكون بشكل مُتَّسِقٍ أقل.

إن أولى الخطوات لإجراء اختبار الوسيط هو دمج كل الدرجات من العينات المنفصلة، وبعد ذلك، إيجاد الوسيط لهذه المجموعات المندمجة. والخطوة الأخرى، هي بناء مصفوفة Matrix تحتوي على عمود لكل عينة من العينات، وصفين، وتحسب درجة للأفراد الذين هم فوق الوسيط، ودرجة للأفراد الذين هم تحت الوسيط. وأخيراً، يتم عد كم عدد الأفراد الذين سجلوا درجات فوق الوسيطات المشتركة Combined Median، وكم عدد الأفراد الذين سجلوا درجات تحت الوسيط. وبعد ذلك تنظم هذه التكرارات المشاهدة في مصفوفة. ويمكن تقييم هذه التكرارات المشاهدة من خلال استخدام مربع كاي للاستقلال.

إن التوزيعات التكرارية وقيمة مربع كاي يمكن حسابهما بالطريقة نفسها التي تم توضيحها في هذا الجزء من هذا الكتاب. إن القيمة الدالة لاختبار مربع كاي تشير إلى أن التعارض بين توزيع أفراد العينة الأولى يكون أكبر مما هو متوقع عن طريق الصدفة. ولتوضيح اختبار الوسيط للعينات المستقلة نورد المثال التالي:

مثال: البيانات التالية تبين درجات تقدير الذات تم الحصول عليها من عينة تتألف من أربعين طفلاً (N=40). وبعد ذلك تم تقسيم هؤلاء الأطفال إلى ثلاث مجموعات منفصلة طبقاً لمستوى الأداء الأكاديمي (عال، متوسط، منخفض). وباستخدام الوسيط يمكن للباحث أن يقيم ما إذا كانت العلاقة بين هذين المتغيرين علاقة دالة.

جدول (18 ـ 10)

درجات تقدير الذات للأطفال وفقاً لثلاثة مستويات من الأداء الأكاديمي [11]

منخفض			متوسط				عال			
	19	11	20	24	13	22		14	22	
- 7	15	13	16	10	22	18	+ 9	18	19	+ 8
+ 3	16	20	19	14	15	19	− 11	21	12	− 2
= 10	18	10	10	11	18	11	= 20	20	18	= 10
	11	15	12	15	19	12		23	20	

المصدر: Ibid , P. 632.

ولأجل اختبار هذه الفرضية ينبغي على الباحث أن يقوم بالخطوات التالية:

- دمج المجموعات الثلاثة لتصبح كما لو كانت هذه المجموعات مجموعة واحدة.

$$N = 40 = 10 + 20 + 10 = N_3 + N_2 + N_1$$

- يتم ترتيب درجات العينة الجديدة المدمجة تصاعدياً أي من أصغر درجة وهي 10 وحتى أكبر درجة وهي 24. ثم يستخرج الوسيط لهذه المجموعة من الدرجات بنفس الإجراءات التي بيناها في أحد مواضع هذا الكتاب.

إن الوسيط لهذه المجموعة المدمجة N = 40 هو M=16+18=34/2=17. أي 20 درجة تماماً فوق هذه القيمة الوسيطية، وعشرون درجة تحت الوسيط. ونجد أن 8 درجات من أصل 10 درجات فوق الوسيط عندما يتعلق الأمر بالأداء الأكاديمي العالي. في حين نجد 9 درجات من أصل 20 درجة فوق الوسيط فيما يتعلق بالأداء المتوسط. بينما نجد أن ثلاث درجات فقط من أصل عشر درجات فوق المتوسط عند الحديث عن الأداء الأكاديمي المنخفض. إن هذه التوزيعات المشاهدة يمكن تنظيمها في المصفوفة التالية:

	الأداء الأكاديمي		
	عال	متوسط	منخفض
فوق الوسيط	8	9	3
تحت الوسيط	2	11	7

والجدول التالي يوضح التوزيعات التكرارية لهذا الاختبار:

	الأداء الأكاديمي		
	عال	متوسط	منخفض
فوق الوسيط	5	10	5
تحت الوسيط	5	10	5

من خلال هذا الجدول يمكننا استخراج قيمة كاي المحسوبة:

$$x^2 = \frac{(8-5)^2}{5} + \frac{(2-5)^2}{5} + \frac{(9-10)^2}{10} + \frac{(11-10)^2}{10} + \frac{(3-5)^2}{5} + \frac{(7-5)^2}{5}$$

$$x^2 = 5.40$$

بعد إيجاد قيمة مربع كاي المحسوبة، نقوم باستخراج قيمة كاي الجَدْوِلِيَّة ومقدارها 5.99 بدرجة حرية (2) وعند مستوى دلالة $\alpha = 0.05$. ولما كانت الدرجة المحسوبة لمربع كاي تساوي 5.40، فإنها لا تقع في المنطقة الحرجة، وعليه لا يمكننا رفض الفرض الصفري (Ho) باعتبار أن هذه البيانات لا تقدم لنا وقائع كافية لنصل إلى نتيجة مفادها أنه لا توجد فروق دالة بين توزيعات تقدير الذات لهذه المجموعات الثلاث من التلاميذ.

وأخيراً نود أن نضمن بعض الملاحظات المتعلقة بتفسير اختيار الوسيط، يمكن إجمالها في الآتي:

أولاً: إن اختبار الوسيط ليس اختباراً للفروق بين المتوسطات، باعتبار أن المتوسط الحسابي لأي توزيع يمكن أن يتأثر بشكل قوي بالدرجات المتطرفة. وعليه، فإن المتوسط والوسيط لأي توزيع ليس بالضرورة أن يكونا متساويين، وأنه ليس من الممكن أن يكونا مرتبطين، إن النتيجة المتحصل عليها من خلال اختبار الوسيط لا نستطيع تفسيرها كما يشار إلى أنه يوجد أو لا يوجد فرق بين المتوسطات.

ثانياً: إن اختبار الوسيط، كما أشرنا، لا يقارن مباشرة الوسيط من عينة واحدة بالوسيط للعينة الأخرى. وبالتالي، فإن اختبار الوسيط ليس اختباراً لدلالة الفروق بين الوسيطات. ولكنه بدلاً من ذلك فهو اختبار يقارن توزيع الدرجات لعينة واحدة في مقابل توزيع الدرجات في عينة أخرى. فإذا كانت العينات قد وزعت بشكل متساوٍ حول نقطة مشتركة (وسيط المجموعة المدمجة)، فإن الاختبار سوف يتضمن أنه لا يوجد فرق ذو دلالة. وعلى الجانب الآخر، فإن إيجاد فرق دال، يشير ببساطة إلى أن العينات ليست موزعة بشكل متساوٍ حول الوسيط المشترك، وبالتالي، فإن أفضل تفسير للنتيجة الدالة هو أنه يوجد فرق في توزيعات العينات(12).

أسئلة للمراجعة:

1- ما هي درجة الحرية للجداول ذات الأبعاد التالية:

– 2×2   – 6×4   4×2   – 3×3

2- ما هي القيم الحرجة لـِ X^2 بدلالة 0.05 و 0.10 = α للجداول ذات الأبعاد التالية:

– 2×4   – 6×4   4×2   – 3×3

3- إذا كان X^2 لعينة يساوي 500 أنتجت قيمة لمربع كاي تساوي 24، هل X^2 بنفس التوزيع ذي الصلة ولكن مع:

N=100 , N=50

4- من الجدول التالي - احسب التوزيعات المتوقعة لكل خلية مبيناً تلك الحالات التي تنتهك القواعد المستخدمة في X^2:

	أ	ب	ج	د	المجموع
أ	1	0	6	48	55
ب	2	0	7	40	49
المجموع	3	0	13	88	104

5- في الفصول السابقة قارنا عينات افتراضية للأطفال من استراليا، كندا، سنغافورة، وبريطانيا، فيما يتعلق بكمية مشاهدة الإذاعة المرئية (التلفزيون). افترض أن هذا المتغير لم يتم قياسه على مستوى المقياس الترتيبي والنسبي، ولكن عوضاً عن ذلك تم قياسه على مستوى المقياس الترتيبي. وقد جاءت نتائج هذا المسح كالتالي:

كمية مشاهدة الإذاعة المرئية	المجتمعات				
	كندا	أستراليا	بريطانيا	سنغافورة	المجموع
منخفضة	23	25	28	28	104
متوسطة	32	34	39	33	138
عالية	28	30	40	35	133
المجموع	83	89	107	96	375

من خلال هذه البيانات، هل يمكنك القول بأن كمية مشاهدة الأطفال للإذاعة المرئية مستقلة عن مكان الإقامة؟

6- عينة مكونة من 162 رجلاً تتراوح أعمارهم بين 40 و 65 سنة، تم سحبها لمعرفة أوضاع هؤلاء الرجال الصحية. تم طرح سؤال على كل رجل لمعرفة فيما إذا كانت عادة التدخين على أساس منظم. وقد تم الحصول على النتائج التالية كما تظهر في جدول التقاطع:

الوضع الصحي	عادة التدخين		
	لا يدخن	يدخن	المجموع
سيء	13	34	47
متوسط	22	19	41
جيد	35	09	44
جيد جداً	27	03	30
المجموع	97	65	162

المطلوب:

بالنظر إلى نسب العمود، هل تعتقد بأن الفروق في مستوى الوضع الصحي لعينة المدخنين وغير المدخنين يمكن إرجاعها إلى تباين المعاينة بدلاً من الفرق في المجتمعات.

- أَجْرِ اختبار مربع كاي للاستقلال لهذه البيانات؟ هل النتيجة التي توصلت إليها تؤكد ما توصلت إليه في الإجابة السابقة؟

7- من البيانات التالية:

الاتجاه

	أوافق	لا أوافق	
ذكور	35	15	50
إناث	55	45	100
	90	60	

أ- بين ما إذا كان هناك فرق ذو دلالة في توزيع الاتجاهات للذكور مقارنة بالإناث.

اختبر هذا الفرض على مستوى دلالة 05.

ب- إن العلاقة بين النوع والاتجاه يمكن الحصول عليها باستخدام معامل ϕ. ما هي قيمة ϕ إذا تم حسابها من هذه البيانات.

8- البيانات السابقة بينت لنا أنه لا يوجد فرق ذو دلالة بين توزيعات الذكور والإناث. ولبناء جدول جديد يمكنك مضاعفة حجم العينة بـِ 15؛ وبالتالي ستكون التوزيعات لكل خلية ضعفين (لاحظ أن تناسب العينة لم يتغير).

الاتجاه

	أوافق	لا أوافق	
ذكور	70	30	100
إناث	110	90	200
	180	120	300

المطلوب:

أ- اختبار فرق الدلالة بين توزيعات الذكور وتوزيعات الإناث مستخدماً مستوى الدلالة 05.، كيف تقارن بين النتائج التي توصلت إليها في التمرين السابق، بما توصلت إليه في التمرين الحالي سوف نجد أنه كلما كبر حجم العينة زاد من احتمالية نتيجة الدلالة.

ب- احسب قيمة ϕ لهذه البيانات، وقارن بين نتيجة ϕ في التمرين السابق، وقيمة ϕ في

التمرين الحالي (نعتقد أنك ستجد أن حجم العينة ليس له أي تأثير على قوة العلاقة).

9- بيانات من قسم المرور بمدينة بنغازي (بيانات افتراضية وليست واقعية) تشير إلى أن 80 % من أولئك الذين يحملون رخص قيادة هم أكبر من 25 سنة. وفي عينة مؤلفة من 50 شخصاً تحصلوا حديثاً على مخالفات مرورية 32 كانوا أكبر من 25 سنة و18 كانوا أعمارهم 25 سنة أو أقل.

السؤال المطروح هو: هل توزيع العمر لهذه العينة يختلف بشكل دال عن التوزيع المتعلق بأولئك الأفراد الذين يحملون رخصاً للقيادة؟ يرجى استخدام 05. = α.

10- دراسة اجتماعية تهدف إلى تقييم ثلاثة أنواع من أجهزة الهواتف الجوالة وأن الباحث الذي يقوم بهذه الدراسة كانت لديه شكوك أن طلاب الجامعة قد تكون لديهم معايير تختلف عن تلك المعايير لدى الجيل الكبير. ولكي يختبر هذه الفرضية. قد أعد الباحث هذه الدراسة مستخدماً عدد 60 فرداً من الجيل الكبير، إضافة إلى عينة مكونة من 60 طالباً والبيانات التالية تبين نتائج هذه الدراسة.

	تصميم (1)	تصميم (2)	تصميم (3)
الطلاب	27	20	13
الجيل الكبير	21	34	5
مج	48	54	18

السؤال: هل هذه البيانات تشير إلى أن أفضلية التوزيع للجيل الكبير تختلف بشكل دال عن التوزيع لدى طلاب الجامعة.. يرجى اختبار هذه الفرضية بمستوى دلالة 05. α.

11- إذا كانت لدينا قيمة فاي تساوي 0.229 + لجدول 2×2 جاءت من عينة قوامها 200. اختبر ما إذا كانت هذه القيمة دالة؟ (اختبر هذه القيمة على مستوى دلالة 0.05 و 0.01).

الهوامش والمصادر:

أولا: الهوامش:

1- J. Richard Kendrick , Social Statistics: An Introduction using SPSS , Second edition , USA , 2005 , P. 353.

2- George Argyrous , Statistics for Social and Health Research , Sage Publications ,London , 2001 , P. 398.

3- Ibid , P. 399.

4- Ibid , P. 403.

5- Ibid , P. 403.

6- Frederick J. Gravetter and Larry B. Wallnau , Statistics for The Behavioral Sciencies , 8th ed , 2010 , PP. 628 - 629.

7- جولي بالانت (ت) خالد العامري، التحليل الإحصائي باستخدام برامج SPSS، دار الفاروق للنشر والتوزيع، مصر، 2009، ص 138.

8- Frederick J. Gravetter and Larry B. Wallnau , op.cit. , PP. 626 - 628.

9- J. Cohen , Statistical Power analysis for behavioral Science , SAN Diego , CA. Academic Press , 1988.

10- Frederick J. Gravetter and Larry B. Wallnau , op.cit. , PP. 629 - 633.

11- Ibid , P. 632.

12- Ibid , P. 633.

ثانيا: المصادر:

1- Frederick J. Gravetter and Larry B. Wallnau , Statistics for The Behavioral Sciencies , 8th ed , Wadsworth Cengate Learing , USA , 2010.

2- George Argyrous , Statistics for Social and Health Research , Sage Publications ,London , 2001.

3- J. Cohen , Statistical Power analysis for behavioral Science , SAN Diego , CA. Academic Press , 1988.

4- J. Richard Kendrick , Social Statistics: An Introduction using SPSS , Second ed , USA , 2005.

5- R. Mark Sirikin , Statistics for Social Science , Sage Publications , International Oaks , London , New Delhi , 1995.

6- جولي بالانت (ت) خالد العامري، التحليل الإحصائي باستخدام برامج SPSS، دار الفاروق للنشر والتوزيع، مصر، 2009 م.

7- عبد الله عامر الهمالي، التقنيات الإحصائية ومناهج البحث، منشورات جامعة قاريونس، 2008 م.

الفصل التاسع عشر
اختبار توزيع ثنائي الحد لعينة واحدة

مقدمة:

تناولنا في الفصول السابقة اختبارات Z و t لمتوسط حسابي لعينة واحدة وقد تم تطبيق الإجراءات الإحصائية لهذين الاختبارين على السؤال البحثي المطروح الذي يوجه استقصاءنا لتوزيعات النزعة المركزية والمتغير الذي نرغب في دراسته والذي تم قياسه على المستويين ذي المسافات والنسبي. ويطلق على مثل هذه الاختبارات، الاختبارات المعلمية باعتبارها تقوم باختبار الفروض حول معلمات مجتمع ما (في هذه الحالة المتوسط).

نشير هنا، إلى أن هناك كثيراً من الأحوال التي نرغب فيها في معرفة مظاهر توزيعات متغير ما أكثر من التوزيعات المتعلقة بالمتوسط الحسابي، كالتوزيع التكراري. ويمكننا حساب العديد من الإحصاءات الوصفية لتلخيص بيانات بحثية، هنا يمكننا القول، بأن هذه الإحصاءات الوصفية ذات العلاقة توظف لمساعدة الباحث في الإجابة على الأسئلة البحثية المطروحة.

عود على بدء إذا ما أخذنا على سبيل المثال، القضية التي تم طرحها سلفاً، وهي السياسة التي انتهجتها الهيئة الوطنية للمعلومات حول مسألة التمويل للمنطقة والذي اعتمد فيها المعيار العمري (المتوسط العمري) لمجتمع يتعدى متوسطه 40 سنة. وبوضوح فإن هذه السياسة هي التي توجه تحليلنا لقيمة المتوسط للمتغير الذي نرغب في دراسته وهو - العمر - ولنفترض جدلاً أن الهيئة العامة للصحة قد غيرت من سياستها وبشكل مفاجئ قررت اعتماد مخصصات إضافية لتلك المنطقة فيما يتعلق بالخدمات الصحية فقط إذا كان 20 % أو أكثر من السكان في هذه المنطقة تتجاوز أعمارهم 40 سنة. وفجْأة أصبح المتوسط الحسابي للعمر ليس له علاقة في اعتماد المخصصات، إلا أننا لازلنا نقوم بحساب المتوسط العمري بالرغم من أن المتوسط لن يساعدنا للوصول إلى اتخاذ قرار حول المخصصات.

إن أفضل طريقة لوصف البيانات هو التعامل مع القاعدة السياسية الجديدة للهيئة وذلك بتقسيم العينة إلى أولئك الناس الذين يبلغون 40 سنة من العمر أو أقل، وأولئك الذين تزيد أعمارهم عن 40 سنة، وبعد ذلك نقوم بحساب نسبة كل فئة من هاتين الفئتين، وكنتيجة لذلك يمكننا أن ننظم بياناتنا في نمط بسيط من التوزيعات التكرارية يطلق عليه "اختبار توزيع ثنائي الحد". فالتوزيع ذو الحدين له قيمتين أو فئتين: مثل النوع يمثل توزيع الذكور والإناث[1].

تجدر الإشارة، إلى أن المتغيرات المقاسة على مستويات: المقياس الترتيبي والمقياس ذي المسافات والنسبي، يمكن اختزالها إلى متغيرات تصنيفية أو ثنائية.

البيانات الاسمية:

إن المتغير الاسمي لا يمتلك في جوهره فئتين فقط في هذه الحالة يمكن اختزاله إلى توزيع ذي حدين من خلال تحديد عدد الحالات التي تقع أو لا تقع داخل فئة بعينها أو مجموعة متألفة من الفئات. فعلى سبيل المثال، إن توزيع الحالات المقاسة على المستوى الاسمي طبقاً للانتماء الديني يمكن تحديدها من خلال خمس فئات:

1- الكاثوليك.

2- البروتستانت.

3- اليهود.

4- الأرثوذكس.

5- الإسلام.

وان هذه الفئات يمكن اختزالها إلى توزيع ثنائي بإحدى الطريقتين:

1- الإشارة إلى نسبة الحالات التي تقع أو لا تقع في أحد هاتين الفئتين الموجودة مثل: الكاثوليك وغير الكاثوليك أو مثل المسلمين وغير المسلمين.

2- من خلال توليد فئتين مختلفتين تماماً، وذلك من خلال توحيد هذه الفئات مثل الكاثوليك وغير الكاثوليك، المسلمون وغير المسلمين، انظر الجدولين التاليين:

جدول (19 ـ 1) الانتماء الديني

الديانة	التوزيعات التكرارية
الكاثوليك	20
البروتستانت	15
الأرثوذكس	12
الإسلام	12
اليهودية	7

الديانة	التوزيعات التكرارية
الكاثوليك	20
البروتستانت	15
الأرثوذكس	14
الإسلام	10
اليهودية	7

جدول (19 ـ 2) الانتماء الديني

الديانة	التوزيعات التكرارية	النسبة
الكاثوليك	20	% 30
غير الكاثوليك	46	% 70

الديانة	التوزيعات التكرارية	النسبة
المسلمون	10	% 15
غير المسلمين	56	% 85

المصدر: George Argyrous, Statistics for Social & Health Research, with a Guide to SPSS , SAGE Publications, London, 2001 , P. 304.

البيانات الترتيبية والبيانات ذات المسافات والنسبية[2]:

البيانات الترتيبية أو البيانات ذات المسافات المتساوية والنسبية يمكن اختزالها إلى توزيعات ثنائية (ذات الحدين) وذلك من خلال تحديد عدد الحالات التي تقع فوق أو تحت قيم محددة على المقياس. على سبيل المثال: إذا كانت لدينا قائمة بدرجات امتحان مادة الإحصاء الاجتماعي عندئذٍ بإمكاننا اختزال هذه الدرجات إلى توزيع ذي حدين، وذلك من خلال اختيار نسبة 50 % كخط التقسيم وتنظيم الدرجات إلى ناجح وراسب.

توزيع المعاينة لنسب عينة[3]:

عندما يُتِمُّ الباحث تنظيم بياناته في شكل توزيع ثنائي ويقوم بحساب النسب المتعلقة بالحالات في كل فئة من هاتين الفئتين، عندئذٍ بإمكانه إجراء اختبار الاستدلال لهذه النسب. ولفعل ذلك، ينبغي علينا معرفة توزيعات المعاينة لنسب العينة.

في الفصول السابقة، كان الاهتمام منصباً على إجراء الاستدلال من متوسط العينة على متوسط المجتمع الذي سحبت منه هذه العينة؛ ولفعل هذا الاستدلال فقد تم بناء توزيعات المعاينة لمتوسطات العينة. ويسمح لنا هذا التوزيع تقييم الاحتمالية في الحصول على متوسط العينة الفعلي من مجتمع بقيمة افتراضية محددة (الفرض الصفري).

تجدر الإشارة، عند العمل بالتوزيع الثنائي فإن الإحصاء الوصفي المحسوب من العينة لم يعد متعلقاً بالمتوسط الحسابي بقدر ما يكون متعلقاً بنسبة الحالات التي تقع داخل إحدى الفئتين المحتملة للمتغير. وبعد حساب نسبة العينة حينئذٍ ينبغي علينا بالضرورة القيام بعملية الاستدلال حول النسبة المتعلقة بالمجتمع ككل. وعليه، فالضرورة تقتضي سبر غور خاصيات توزيعات المعاينة لنسب العينة: إن توزيعات نسب المعاينة سوف تظهر من خلال إعادة تكرار العينات العشوائية ذات الأحجام المتساوية. على سبيل المثال، يمكننا معرفة أن 50 % من كل الطلاب في الجامعة (افتراضية) هم من الذكور وأن 50 % من الإناث وبالرغم من هذا، إذا ما تم سحب عينة من 100 طالب من طلاب الجامعة فإنه ليس بالضرورة أن نحصل من خلال هذه العينة على 50 % من الذكور و 50 % من الإناث، فقد يكون التباين العشوائي مسئولاً عن بعض العينات التي تحتوي على عدد

أكبر قليلاً من الإناث، بينما في عينات أخرى سوف تحتوي على عدد أكبر قليلاً من الذكور. لكنه يمكننا القول، بأن معظم هذه العينات المكررة سوف تحتوي على نسبة كل نوع إما مساوٍ أو قريب من 50 %، بمعنى آخر، بينما يوجد بعض التباين في التوزيعات المكررة لنسب العينة، فإن هذه النسب سوف تتجمع حول القيمة الحقيقية للمجتمع وهي 50 %. وإذا سحبنا عدداً لا متناهياً من العينات العشوائية بأحجام متساوية من مجتمع، وقمنا بحساب نسبة الحالات في كل حالة تمتلك قيمة محددة للتوزيع ذي الحدين، فإن توزيع المعاينة لنسب هذه العينات سوف تحتوي على الخاصيات التالية:

1- إن توزيع المعاينة سيكون قريباً من التوزيع المعتدل بنسبة وسيط مساوٍ لقيمة المجتمع. فتوزيع المعاينة يكون فقط، قريباً من المعتدل، لأن التوزيع الثنائي هو متغير منفصل. ولما كان المنحنى الطبيعي هو منحنى متصل، ومع ذلك كلما كان حجم العينة كبيراً، كانت التوزيعات تقارب التوزيع المعتدل.

2- إن الخطأ المعياري لتوزيع المعاينة يمكن تعريفه بالمعادلة التالية:

$$\sigma p = \sqrt{\mathbf{pu}\left(\frac{\mathbf{1 - Pu}}{\mathbf{N}}\right)}$$

حيث إن: Pu تساوي نسبة المجتمع.

تعتبر هذه المعلومات المتعلقة بهاتين النقطتين السابقتين في غاية الأهمية كما تبين لنا في الفصول السابقة.

إن معرفتنا بكل الاحتمالات الممكنة لنسب العينة التي جاءت من مجتمع معين تسمح لنا بحساب الاحتمالية في الحصول على نتيجة عينة محددة من مجتمع بقيمة افتراضية، على سبيل المثال، إذا كانت عينة ما تحتوي على 60 % من الإناث، فإنه باستطاعتنا حساب الاحتمالية بأن هذه النتيجة كانت نتيجة خطأ المعاينة عند عملية سحب العينة من المجتمع الذي تصل فيه نسبة الإناث إلى 50 % فقط. إنه بالتحديد ذلك النوع من السؤال الذي سنتعامل معه، صمم اختبار Z لنسب العينة.

اختبار لنسبة توزيع اختبار ثنائي الحد[4]:

بالرغم من وصفنا للبيانات من خلال تنظيمها وفقاً للتوزيع ثنائي الحد بدلاً من حساب المتوسط الحسابي لهذه البيانات فالإجراءات المتبعة في الاستدلال من العينة على المجتمع هي إجراءات مماثلة. فمن الناحية العملية، فإن الخطوات المتعلقة باختبار الفرض للنسب هي نفسها التي تم اتباعها في اختبار الفرض للمتوسط الحسابي.

لقد أجرينا الاختبار الاستدلالي كما تم إجراؤه في الفصلين المتعلقين باختبار Z واختبار t لعينة واحدة. إلا أننا من خلال هذا الفصل، فإن اختبار الاستدلال سيعتمد على نسبة العينة الواقعة في واحدة من الفئتين من التوزيع الثنائي، بدلاً من الاعتماد على المتوسط الحسابي للعينة. وبما أن توزيعات المعاينة هي توزيعات طبيعية، فإننا سنجري اختبار Z للفرق بين نسب العينة والقيمة المفترضة (هو إجراء مشابه لما قمنا به عند اختبار Z للفرق بين متوسط العينة ومتوسط المجتمع المفترض).

إن المعادلات المحددة لحساب Z للنسب تكون كالتالي:

$$Z = \frac{(Ps - 0.5) - Pu}{\sqrt{\frac{Pu(100 - Pu)}{N}}} \quad \text{العينة}$$

حيث إن: $Ps > Pu$

أو:

$$Z = \frac{(Ps + 0.5) - Pu}{\sqrt{\frac{Pu(100 - Pu)}{N}}} \quad \text{العينة}$$

حيث إن: $Ps < Pu$

Ps نسبة العينة.

Pu نسبة المجتمع.

لاحظ أن إشارة الجمع والطرح لنسبة 0.5 إلى أو من نسبة العينة في هاتين المعادلتين قد حُدِّدَتْ بشكل صارم لأن التوزيع ذا الحدين ليس موزعاً توزيعاً طبيعياً، وبالتالي فإن

عملية الجمع أو الطرح لـِ 0.5 (التصحيح المتواصل) تمدنا بأفضل قيمة تقريبية للتوزيع الطبيعي. ففي العينة التي تزيد عن 30 مفردة فإن هذا التقريب سيكون مناسباً ودقيقاً. أما على الجانب الآخر، إذا كانت العينة أقل من 30 مفردة، فإن التقريب لا يكون دقيقاً، وبالتالي، فإن اختيار احتمالية دقيقة لتوزيع ثنائي يمكن استخدامه.

مثال: نفترض أننا نرغب في معرفة معدل البطالة في منطقة معينة أصابها ركود اقتصادي صعب، مقارنة بباقي أجزاء البلاد. وأن الباحث على معرفة بأن معدل البطالة على مستوى البلاد ككل يصل إلى 11 % وقد قرر سحب عينة من 120 فرداً. وطرح عليهم السؤال التالي: ما إذا كان هؤلاء الأفراد خارج سوق العمل أم لا؟ وقد كانت النتائج المستخلصة من هذه العينة، أن 18 فرداً من إجمالي العينة أقروا بأنهم عاطلون عن العمل. وقد فرغت إجابات هؤلاء المبحوثين على السؤال البحثي المطروح في توزيع ثنائي كما هو موضح في الجدول التالي:

جدول (19 ـ 3): توزيع المبحوثين حسب الوضع المهني

الوضع المهني	التكرار	النسبة المئوية
يعمل	102	% 85.0
لا يعمل (عاطل عن العمل)	18	% 15.0
المجموع	120	% 100.0

المصدر: George Argyrous, Statistics for Social & Health Research, OP. Cit. , P. 307.

هل هذه البيانات حقاً تشير إلى أن هذه المنطقة قد تعرضت لركود اقتصادي صعب؟

الخطوة الأولى: صياغة الفرض الصفري والفرض البديل:

Ho الفرض الصفري: إن نسبة البطالة في هذه المنطقة المحلية مساوية لنسبة البطالة في باقي البلاد.

$$Ho = P_u = \% \ 11$$

Hi الفرض البديل: إن المنطقة المحلية لديها أعلى نسبة من الأفراد العاطلين عن العمل عن باقي البلاد.

$$Hi = P_u > \% \ 11$$

الخطوة الثانية: اختيار درجة الدلالة:

إن السؤال البحثي المرغوب فيه، هو نسبة الناس في فئة التوزيع الثنائي (ذي الحدين) على سبيل المثال الأفراد العاطلين عن العمل. وعليه فإننا في هذه الحالة سنستخدم اختبار Z أحادي الجانب للنسب.

الخطوة الثالثة: حساب درجة العينة:

إن بيانات هذه العينة المناسبة من التوزيع الثنائي أعلاه تكون:

$$N= 120$$

$$Ps = \frac{18}{120} \times 100 = 15\%$$

بالتعويض نتحصل على:

العينة $$Z = \frac{(Ps - 0.5) - Pu}{\sqrt{\frac{Pu(100 - Pu)}{N}}} = \frac{(15 - 0.5) - 11}{\sqrt{\frac{11(100 - 11)}{120}}}$$

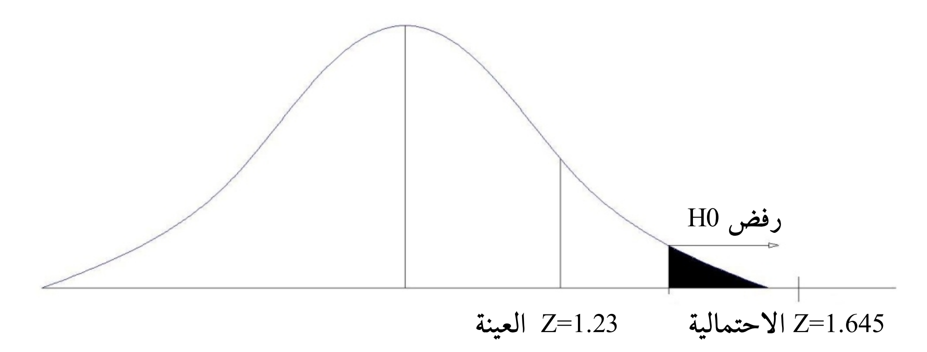
$$= 1.23$$

الخطوة الرابعة: اختيار الدرجات الحرجة والمنطقة الحرجة:

لقد تم اعتماد مستوى الدلالة لهذا المثال بـِ 0.05 (∝) ولشك الباحث أن هذه المنطقة المحلية قد تعرضت لركود اقتصادي صعب مما حدا به لاستخدام اختبارٍ أحاديٍّ الجانبِ. ولما كان اتجاه الفرق قد حدد من خلال الفرض البديل بأن معدل البطالة في هذا المجتمع المحلي يكون أعلى من المعدل الوطني العام. وعليه فإن الجانب المناسب من توزيع المعاينة لهذا الاختبار هو الجانب الأيمن. ومن خلال استخدام الجدول تحت المنحنى الطبيعي المعياري، فإن درجة Z الحرجة هي 1.645 +.

الخطوة الخامسة: اتخاذ القرار:

إن قيمة Z المحسوبة تساوي 1.23 وهي بالتالي أقل من درجة Z الاحتمالية 1.645، انظر الشكل التالي:



شكل (19 ـ 1) درجات العينة والمنطقة الحرجة

بالرغم من أن معدل البطالة في العينة أكثر منه على المستوى الوطني، إلا أن الفرق بين هذين المعدلين ليس كبيراً يجعلنا نقر بأن السكان في هذه المنطقة يكابدون معدلاً عالياً من البطالة. وعليه لا يمكننا رفض الفرض الصفري على مستوى دلالة 0.05.

إجراء اختبار ثنائي الحد Binomial Test باستخدام SPSS:

1- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Analyze ⇐ Nonparametric Test ⇐ Binomial

2- من قائمة المتغير انقر على Employment Status .

3- انقر على ◄ يقوم بلصق Employment Status في Test Variable List.

4- في المربع المقابل لِـ Test Proportion تطبع 11.

5- انقر على ok.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

Npar Tests Binomial Test

	Category	N	Observed Prop.	Test Prop.	Asymp. Sig. One taild
Employment Status Group 1	UN Employed	18	.15	.11	.105[a]
Group 2	Employed	102	.85		
TOTAL		120	1.00		

a Based ON Z approximation

المصدر: George Argyrous, Statistics for Social & Health Research, with a Guide to SPSS , SAGE Publications, London , 2001 , p. 309.

شكل رقم (19 ـ 1) مخرجات SPSS لاختبار Z الثنائي

تفسير مخرجات SPSS لاختبارات عينة واحدة لتوزيع ثنائي الحد:

من خلال المربع أعلاه يتبين لنا أن مخرجات SPSS أنتجت لنا الاختبار في شكل تناسب بدلاً من النسب. وهذه عملية لا تؤدي إلى اختلاف جوهري حيث إنه يمكن تحويل التناسب إلى نسب وذلك من خلال تحريك الفاصلة العشرية Decimal Point مكانين إلى اليمين.

في مربع اختبار التوزيع ثنائي الحد Binomial test لدينا عمود يحتوي على تكرار الحالات لكل فئة من فئات التوزيع ثنائي الحد. وكذلك عمود يشير إلى التوزيعات النسبية Relative Frequencies كتناسب Proportions، والعمود الأخير يشير إلى الدلالة المقاربة Asym. Sig.(1-taild) اختبار أحادي الجانب وهو عمود مهم لغرض الاختبار الاستدلالي. وبالرغم من أنه لن يعطي قيمة Z (العينة) فقد أعطيت احتمالية أحادي الجانب المرتبطة به. وهنا احتمالية أحادي الجانب 105. التي تشير إلى أنه إذا كان الفرض الصفري فرضاً صحيحاً على الأقل 1 في 10 عينات ستكون معدلات البطالة 15 % أو أكثر. إن الادعاء أن الفرض الصفري فرضُ صحيحُ لا نستطيع رفضه.

تجدر الإشارة إلى أن مخرجات SPSS دائماً تعطينا اختباراً أحادي الجانب عندما تحدد نسبة الاختبار بدلاً من عدم وجود قيمة 0.05. وإذا كان الفرض البديل يتطلب

اختباراً ثنائي الجانب at two-tail test عندئذ وبكل بساطة تضاعف احتمالية اختبار أحادي الجانب. وعلى سبيل المثال إذا كان توزيع المعاينة لاختبار أحادي الجانب لدرجة Z تحتوي على 0.105 من المسافة تحت المنحنى الطبيعي، فإن اختبار ثنائي الجانب يصل إلى 0.21 من المسافة تحت المنحنى الطبيعي[5].

تقدير نسبة مجتمع[6] :

تقدير نسبة مجتمع تعتبر عملية شائعة الاستخدام في مسوح الرأي العام وعادة ما نقرأ في أحوال كثيرة في الصحف أن نسبة معينة من المجتمع تفضل هذه القضية أو تلك، وعادةً لا تمثل هذه النسب كل أفراد المجتمع ولكن من خلال عينة يتم سحبها لهذا الفرض، الأمر الذي يتطلب منا تقدير قيمة المجتمع من نتائج العينة المسحوبة من هذا المجتمع وتستخدم المعادلة التالية لبناء فترة الثقة للنسب.

$$Ci = Ps \pm Z\sqrt{\frac{Ps(100 - Ps)}{N}}$$

إذاً يمكن استخدام هذه المعادلة لبناء فترات الثقة لتقدير نسبة الناس الذين يُدْلُونَ – على سبيل المثال – بأصواتهم لأحد الأحزاب المتصارعة على السلطة، واستناداً على بيانات العينة التالية يمكننا بناء فترات الثقة هذه.

$$N = 120$$

$$Ps = \%\ 15$$

$$\alpha = 0.05 ,\ Z = (1.96)$$

من هذه البيانات يمكننا حساب فترة الثقة:

$$Ci = Ps \pm Z\sqrt{\frac{Ps(100 - Ps)}{N}} = 15 \pm 1.96\sqrt{\frac{15(100 - 15)}{120}}$$

$$= 15 \pm 6.4$$

وهكذا يكون الحد الأدنى لفترة الثقة 8.6 في المائة (6.4 – 15) والحد الأعلى لفترة الثقة 21.4 في المائة (6.4 + 15).

مثال إضافي لزيادة التوضيح[7]:

عينة من 500 طالب كانوا قد سجلوا في مسار تحليل البيانات بالدراسات العليا مقابل رسوم دراسية. منهم 55 % ينحدرون من أسر ذات خلفية اجتماعية - اقتصادية متدنية. ما هي نسب التقدير لكل طلاب الدراسات العليا الذين قاموا بدفع الرسوم الدراسية مقابل هذا المسار، وينحدرون من أسر ذات وضع اقتصادي واجتماعي متدنٍّ؟

$$n = 500$$

$$Ps = \frac{55}{500} \times 100 = \%11$$

$$\alpha = 0.05\ ,\ Z = (1.96)$$

$$Ci = Ps \pm Z\sqrt{\frac{Ps(100 - Ps)}{N}}$$

$$= 11 \pm 1.96\sqrt{\frac{11(100 - 11)}{500}}$$

$$= 11 \pm 2.7$$

بمعنى آخر، إن نسبة الطلاب ذوي الخلفية الاقتصادية والاجتماعية المتدنية في السكان تقع بين (8.3 % و 13.7 %) بمستوى ثقة (95 %).

الاستدلال باستخدام فترة الثقة للنسب[8] :

إن السبب وراء إجراء المسح السابق هو معرفة ما إذا كان اعتماد الرسوم الدراسية لمسارات برنامج الدراسات العليا يؤدي إلى نتيجة معاكسة على المجموعات المحرومة من الطلاب.

إن الإحصاءات السكانية المتوفرة لطلاب الدراسات العليا قبل إدخال اعتماد الرسوم الدراسية الجامعية، كانت تشير إلى 16.8 % من طلاب الدراسات العليا ينحدرون من أسر ذات وضع اقتصادي واجتماعي متدنٍ. هل يمكننا القول، بأن إدخال الرسوم الدراسية قد أثرت على دخول الطلاب الفقراء للجامعة، نريد أن نقارن عدد طلاب الدراسات العليا قبل إدخال الرسوم الدراسية بعدد الطلاب بعد الإدخال. ونود

أن نشير إلى أنه لدينا معلومات قبلية فيما يتعلق بالرسوم الطلابية ولكن لدينا فقط تقديرات بعدية للرسوم المتعلقة بطلاب الدراسات العليا استناداً على عينة، وعلى أية حال يمكننا أن ندرك تقدير نسبة المجتمع للرسوم البعدية والتي لا تتضمن 16.8 %. إن نسبة الطلاب المحرومين فيما يتعلق بالرسوم القبلية لا تقع ضمن 95 % فترة ثقة ضمن ما قدرناه. حيث تكمن في نسبة الطلاب بعد إكمال الرسوم الدراسية البعدية لدى الطلاب الذين جاءوا من خلفيات اجتماعية واقتصادية متدنية، بمعنى آخر، استناداً على نتائج العينة يمكننا القول بأن نسبة الطلاب المحرومين في فترة قبل اعتماد الرسوم الدراسية كانت نسبة عالية وذات دلالة أكثر منها بعد إدخال الرسوم الدراسية. وعليه، يمكننا القول، بأن إدخال الرسوم الدراسية كان له تأثير على المجموعة المحرومة من الطلاب.

عليه، يمكننا إجراء اختبار الاستدلال لمعرفة ما إذا كنا قادرين على رفض الفرض الصفري حول نسبة مجتمع من خلال بناء فترة ثقة ويعتبر هذا بديلاً لاختبار الفرض الذي تعلمناه من خلال الفصول السابقة. في الحقيقة سوف نقوم بإجراء اختبار Z لعينة واحدة للنسب، وللوقوف على هذه النقطة بشكل أكبر، لنبيّن أن النتيجة ستكون واحدة عند استخدام فترة الثقة للوصول إلى قرار حول الفرض:

العينة

$$Z = \frac{(Ps - 05) - Pu}{\sqrt{\frac{Pu(100 - Pu)}{N}}}$$

$$= \frac{(11 - 05) - 16.8}{\sqrt{\frac{(16.8)(100 - 16.8)}{500}}}$$

$$= 3.8$$

في ضوء اختبار ثنائي الجانب، فإن قيمة Z الحرجة بمستوى دلالة 0.05 (مساوية لـ 95 % مستوى ثقة) ستكون ±1.96 وبوضوح، فإنه بإمكاننا أن نرفض الفرض الذي مفاده أن هذه العينة جاءت من مجتمع يمثل نسبة المحرومين فيه من الطلاب 16.8 %.

العلاقة بين مربع كاي واختبار Z ثنائي الحد:

ناقشنا في الفصل السابق بشكل واسع اختبار مربع كاي للاستقلال، باعتباره من التقنيات الإحصائية لاختبار الدلالة الأكثر شيوعاً واستخداماً في مجال البحث الاجتماعي، ويعزى سبب شيوع استخدام مربع كاي أنه يمكن تطبيقه في مواقف يكون فيها لدينا بيانات مصنفة تصنيفاً اسمياً وترتيبياً، ناهيك عن رغبتنا في الحصول على التوزيع التكراري عبر فئات المتغير الخاضع للدراسة. ويعتبر هذا الموقف موقفاً شائعاً في البحث الاجتماعي.

إن اختبار مربع كاي يفحص توزيع الإجابات في جدول ثنائي يقيم ما إذا كان نمط التبعية موجوداً في حالة الجدول الثنائي 2 × 2 (عندما يكون كلا المتغيرين ثنائي الحد binominal). فإن اختبار Z للنسب يمكن إجراؤه على البيانات نفسها؛ وفي الحقيقة أن كلا الاختبارين متكافئُ في طرق تحليل نفس البيانات، والوصول إلى النتيجة نفسها أي $(X^2 = Z^2)$.

حقاً إن اختبار Z للنسب يمكن اعتباره حالة خاصة لاختبار مربع كاي. ولما كان اختبار مربع كاي اختباراً شائع الاستخدام في البحوث، أصبح من الأهمية بمكان معرفة آليات وطرق حسابه.

في هذه الجزئية من هذا الفصل سوف نتعامل مع مثال لاختبار Z لنسب عينة، وبعد ذلك نجري اختبار مربع كاي لنبرهن على أن النتائج التي نتوصل إليها ستكون واحدة وأن أي منهما يمكن استخدامه.

مثال تطبيقي:

لقد تم إجراء مسح (افتراضي) لاستقصاء مستوى الدعم لإصلاح الرعاية الاجتماعية، وما إذا كان هناك تفاوت Varies حسب العمر في هذا المسح. لقد تم ترتيب المبحوثين طبقاً للفئة العمرية: تحت 45 سنة أو 45 سنة أو أكثر. وقد تم طرح سؤال على كل مبحوث مفاده ما إذا كان ينبغي على الدولة أن تقوم بجهد أكبر للتخفيف من الفقر. وقد تم وضع هذا السؤال بصيغة مبسطة (نعم – لا). وقد تمت صياغة الفرض الصفري بأن

نسبة الذين أجابوا بنعم تحت 45 سنة (P_1) تكون مساوية لنسبة أولئك الذين أجابوا بنعم 45 سنة أو أكثر (P_2):

$$Ho: P_1 = P_2$$

إذا كانت هذه الفرضية صحيحة، فإن العينات المسحوبة من مثل هذه المجتمعات ففي العادة سوف تعكس صورة متساوية. بمعنى آخر، أن الفرق بين أي نسب للعينتين، إذا لم يكن هناك فرق بين المجتمعات ينبغي أن يكون هذا الفرق صفراً أو قريباً منه.

تجدر الإشارة إلى أن هذا الأمر لن يكون دائماً على هذا النحو. فالعينات ليست دائماً تعكس خصائص المجتمع التي سحبت منها. فالتباين العشوائي قد يقودنا إلى أن نختار زيادة إضافية طفيفة من صغار السن الذين يؤيدون إصلاح نظام الرعاية الاجتماعية، وكذلك نختار مجموعة إضافية طفيفة من كبار السن الذين يعارضون هذا الإصلاح.

إن مثل هذه الاختبارات الإضافية الطفيفة تسبب إلى حد كبير اختلافاً في نسب العينة. وهذا يعني أنه إذا كان هناك فرق بين نسب العينتين، فإننا لا نستطيع بشكل تلقائي أن نقرر أن هذا الفرق يعكس ما يبطن من فرق في المجتمعات. وعليه فإن الفروق الكبيرة بين نسب العينات يكون أقل احتمالاً بأن تعزو هذا الفرق إلى الفرضية العشوائية. إن اختبار Z للنسب يعطي احتمالية دقيقة أن مثل هذه الوقائع احتمالية حدوثها تكون بعيدة.

إن المسح الذي يحتوي على 600 شخص تحت الفئة العمرية 45 سنة و 400 شخص أعمارهم 45 سنة أو أكثر. إن النسبة لكل مجموعة أجابت "بنعم" (أي يتوجب على الدولة بذل جهد كبير للتخفيف من الفقر) جاءت كالتالي:

-1 تحت سن 45 سنة : $P_1 = \frac{490}{600} \times 100 = 82\%$

$N_1 = 600$

-2 45 سنة أو أكبر : $P_2 = \frac{232}{400} \times 100 = 58\% : 45$

$N_2 = 400$

السؤال الذي ينبغي طرحه في هذا السياق هو هل هذه النسب تعكس ما يتضمنه الفرق بين المجموعات العمرية في هذه القضية المطروحة؟ لتحديد ذلك نبدأ بالمعادلة التالية:

$$Pu = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{600(82) + 400(58)}{600 + 400}$$
$$= 72.4\%$$

إن هذه النتيجة التي توصلنا إليها تسمح لنا بأن نحدد الخطأ المعياري Standard error لتوزيع المعاينة لكل فروق العينة المحتملة. وإن خطأً معيارياً واحداً يمكن تعريفه بـِ:

$$\sigma P - P = \sqrt{Pu(100 - Pu)}\sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$
$$= \sqrt{72.2(100 - 72.2)}\sqrt{\frac{600 + 400}{600(400)}}$$
$$= 2.9$$

إن الفرق الحقيقي بين هاتين العينتين فيما يتعلق بدرجة Z هو:

$$Z = \frac{p_1 - p_2}{\sigma P - P} = \frac{82 - 58}{2.9} = 8.3$$

إن درجة Z هذه تعتبر ذات دلالة على مستوى ألفا $\alpha = 0.01$ وعليه نرفض الفرض الصفري H0 الذي مفاده لا فرق $P_1 = P_2$ وأنه يمكننا المجادلة بأن دعم الدولة لمساعدة الفقراء يتباين وفقاً للعمر [9].

إن الطريقة البديلة لتحليل هذه البيانات هو أن يتم تنظيمها في جدول ثنائي 2×2 كما هو موضح أدناه:

جدول (19ـ4): الاتجاه نحو سياسة الدولة نحو إصلاح نظام الرعاية الاجتماعية حسب مجموعة العمر

الاتجاه	مجموعة العمر		
	تحت 45 سنة	45 سنة أو أكثر	المجموع
لا	(166.8) 110	(111.2) 168	(27.8%) 278
نعم	(433.2) 490	(288.8) 232	(72.4%) 722
المجموع	600	400	1000

المصدر: George Argyrous, Statistics for Social & Health Research,op.cit, , p.417.

تشير الأرقام بين الأقواس إلى القيم المتوقعة وذلك استناداً إلى النسبة الكلية للمبحوثين الذين أجابوا "بنعم" أو "لا"، مع ملاحظة أن نسبة 72.4 في المائة من كل المبحوثين قد وافقوا على ضرورة إصلاح نظام الرعاية الاجتماعية. ومن خلال هذا العدد يمكننا حساب عدد المبحوثين تحت 45 سنة وكذلك حساب عدد المبحوثين ذوي الأعمار 45 سنة أو أكثر الذين يتوقع موافقتهم على إصلاح نظام الرعاية الاجتماعية. إن نسبة 72.4 في المائة هي النسبة ذاتها التي وردت في اختبار Z للنسب لعينتين كنقطة مرجعية لحساب الانحراف المعياري لتوزيع المعاينة. في الحقيقة، أنه عندما نقوم بحساب مربع كاي:

$$X^2 = \sum \frac{(Fo - Fe)^2}{Fe}$$

$$= \frac{(110-168.8)^2}{168.8} + \frac{(168-111.2)^2}{111.2} + \frac{(490-433.2)^2}{433.2} + \frac{(323-288.8)^2}{288.8}$$

$$X^2 = 68 = Z^2 = 8.3$$

من الجدول لتوزيع مربع كاي، فإن الاحتمالية للحصول على هذه القيمة (الأكبر) من مجتمعات مماثلة هي 0.005 – وهي مساوية لتلك الاحتمالية المتعلقة باختبار Z.

إن النتيجة التي تستنتج من هذا هي أنه بينما تكون اختبارات Z لعينتين ثنائية الحد، اختبارات شائعة جداً، وعليه، أصبح الإلمام والمعرفة بهذا الاختبار مهماً جداً، وفي الحقيقة

فإن هذه الاختبارات تمثل حالة خاصة لمربع كاي. وبما أن معادلة اختبار Z هي أكثر إرهاقاً وأن منطقها غير واضح إدراكياً، وبالتالي فإن استخدام اختبار مربع كاي على الأرجح سيكون مناسباً في معظم المواقف[10].

أسئلة للمراجعة:

1- من مجموع البيانات التالية أجرِ اختبار Z للنسب مستخدماً كل من: اختبار أحادي الجانب واختبار ثنائي الجانب بألفا (∝) 0.05.

A	B
$P_u = 52$	$P_u = 42$
$P_s = 61$	$P_s = 39$
$n = 110$	$n = 110$

2- عينة من 900 سجين تم مسحها لضمان نجاح برنامج إعادة التأهيل بالسجن. 350 من هؤلاء السجناء أوضحوا بأن البرنامج كان فعالاً للتقليل من احتمالية إعادة ارتكاب الجريمة. لقد كان الهدف من هذا البرنامج هو تحقيق 40 % من معدل النجاح في التقليل من احتمالية العود إلى ارتكاب الجريمة:

أ- باستخدام اختبار Z للنسب، هل يمكننا القول بأن هذا البرنامج قد نجح.

ب- بناء 95 % فترة ثقة لتقدير قيمة المجتمع. كيف يمكن أن تؤكد فترة الثقة نتيجة اختبار Z.

3- أجريت دراسة على 500 شخص. 56 % من هؤلاء يؤيدون تولي المرأة مناصب عليا في الإدارة. ما هي فترة الثقة لـِ 95 % لنسبة كل الناس الذين يؤيدون تولي المرأة مناصب عليا في الإدارة؟ هل يمكنك القول بأن أغلبية الناس يؤيدون هذا الاتجاه؟

4- عينة عشوائية لعدد 60 مصنعاً. وجد منها 15 مصنعاً لا تتقيد بمعايير السلامة البيئية. ما هي فترات الثقة لـِ:

a) 90 %.

b) 95 %.

الهوامش والمصادر:

أولا الهوامش:

1- George Argyrous, Statistics for Social & Health Research, with a Guide to SPSS , SAGE Publications, London, 2001, P. 303.

2- Ibid. , P. 304.

3- Ibid. , PP. 305 - 306.

4- Ibid. , P. 308.

5- Ibid. , P. 310.

6- Ibid. , PP. 311 - 312.

7- Ibid. , P. 312.

8- Ibid. , PP. 312 - 318.

9- Ibid. , P. 417.

10- Ibid. , P. 418.

ثانيا: المصادر:

1- George Argyrous, Statistics for Social & Health Research, with a Guide to SPSS , SAGE Publications, London, 2001.

2- Frederick J. Gravetter and Larry B. Wallnau , Statistics for The Behavioral Sciencies , Wadsworth Cengage Learning , USA , 2010.

3- Joseph F. Healey , The Essentials of Statistics: A Tool for Social Research , Wadsworth Cengage Learning , USA , 2010.

الفصل العشرون

الأساليب الإحصائية للبيانات الترتيبية

اختبار الفرض:

في هذا الفصل سوف نتناول الإجراءات المتعلقة باختبار الفرض من البيانات مقاسة علي المستوى الترتيبي. وأن كل واحد من هذه الاختبارات يمكن النظر إليه كبديل للاختبارات المعلمية التي تناولناه في الفصول السابقة، وأن الاختبارات الأربعة التي سنتطرق لها، والمواقف التي تستخدم فيها هي كالتالي:

1- اختبار مان- وتني Mann-Whitney

يستخدم هذا الاختبار بيانات من عينتين منفصلتين وذلك لتقييم الفرق بين مجتمعَيْن. ومن هنا يمكن اعتبار اختبار مان- وتني اختباراً بديلاً لاختبار t لعينتين مستقلتين.

2- اختبار ولكوكسن Wilcoxon Test

لقد صُمِّمَ هذا المقياس لتقييم الفرق بين معالجتين. وذلك باستخدام بيانات تم الحصول عليها من تصميم متكرر القياسات a repeated-measures، ويعتبر هذا الاختبار اختباراً بديلاً لاختبار t لمقاييس متكررة a repeated t-test.

3- اختبار كروسكال-ويلز Kruskal-Wallis Test

يُسْتَخْدِمْ هذا الاختبار بياناتٍ لتقييم الفروق بين ثلاث عينات أو أكثر مستخدماً بيانات تم الحصول عليها من خلال تصميم المقاييس المستقلة. ويعتبر هذا المقياس مقياساً بديلاً لتحليل التباين أحادي الجانب Anova.

4- اختبار فريدمان The Friedman test

يستخدم هذا الاختبار بيانات تم الحصول عليها من تصميم متكرر القياسات a repeated-measurea design، وذلك لمقارنة الفروق بين ثلاثة أو أربعة أحوال من المعالجة. ويعتبر هذا الاختبار اختباراً بديلاً لتحليل أنوفا ANOVA لمقاييس متكررة.

وتجدر الإشارة إلى أن هذه الاختبارات تكون ملائمة للاستخدام عندما لا يكون في مقدور الباحث أن يفي بمتطلبات الاختبارات البارامترية (المعلمية).

وبشكل عام، إذا كانت البيانات المتحصل عليها بيانات ملائمة لإجراء اختبار Anova أو أي من اختبارات t. حينئذٍ يكون الاختبار المعياري standard test مفضلاً للبيانات الترتيبية البديلة [1].

5- الحصول علي المقاييس الترتيبية:

للحصول علي الرتب من الملاحظة المباشرة، يمكن للباحث أن يبدأ بمجموعة من المقاييس العددية وتحويل هذه الدرجات إلى رتب Ranks. فعلى سبيل المثال، إذا كان لدى الباحث الأطوال الحقيقية لمجموعة من الأفراد، فإنه بإمكانه أن يرتب الأعداد من الأكبر إلي الأصغر. ويسمح هذا الإجراء بتحويل البيانات من بيانات مقاسة على المستوى ذي المسافات والنسبي إلى مقاييس ترتيبية. إن القائمة التالية سوف تلقي الضوء على المميزات وراء استخدام الرتب بدلاً من الدرجات.

1- إن عملية الرتب هي عملية بسيطة. فإذا سألك أحد عن كم طول ابنك، فإنك ببساطة ستجيبه بقيمة عددية محددة مثل: 170 سم، أو 164سم..... إلخ. أو أنك قد تجيبه بأنه أقصر منك بقليل أو يزيد طوله قليلاً عن طولك.

2- إن الدرجات الأصلية يمكن أن تنتهك بعض الافتراضات الأساسية التي تتضمنها بعض الإجراءات الإحصائية. على سبيل المثال، إن اختبارات t واختبارات Anova تفترض أن البيانات قد جاءت من توزيعات طبيعية Normal distributions، أيضاً فإن اختبارات المقاييس المستقلة Independent measures، تفترض أن مجتمعات مختلفة لديها نفس التباين (فرضية تجانس التباين). أما إذا شك الباحث في أن البيانات لا تفي بهذه الافتراضات، عندئذٍ ينبغي عليه تحويل هذه الدرجات إلى رتب، واستخدام التقنيات الإحصائية المصممة للرتب.

3- إن الدرجات الأصلية قد يكون لديها تباين عالٍ؛ ويعتبر هذا التباين المكون الأساسي للخطأ المعياري في المقام المتعلق لمعادلة إحصاء t. ومصطلح الخطأ في مقام معادلة اختبار f. وعليه، فإن التباين الكبير يمكن من خلاله بشكل واضح تقليل احتمالية أن هذه الاختبارات البارامترية (المعلمية) سوف تكتشف لنا فروقاً دالة.

إن تحويل الدرجات إلى رتب، سيقلل بشكل جوهري من عملية التباين، على سبيل المثال، إذا تم ترتيب عشر درجات (10) من 1 إلى 10، فإن الأمر لا يهم كم كانت الدرجات الأصلية للمتغير.

6- ترتيب الدرجات المتعادلة Ranking Tied Scores:

كلما قام الباحث بتحويل الدرجات العددية إلى رتب، فإنه بالتالي قد يجد رقمين أو أكثر يحملان تماماً نفس القيمة. ولأن الدرجات كانت متعادلة، فإن عملية إجراء التحويل يجب أن تولد رتباً تكون أيضاً متعادلة.

إن عملية إجراء تحويل الدرجات المتعادلة إلى رتب، قد تمت مناقشتها عند الحديث عن معامل سبيرمان للرتب. ولكننا قد نعيد هذا الإجراء بشكل مختصر في هذا الفصل. أولاً ينبغي على الباحث أن يقوم بتنظيم الدرجات في قائمة تشتمل على القيم المتعادلة. وثانياً يحدد لكل حالة في القائمة رتبة معينة (الأول، الثاني..... إلخ). ثالثاً وأخيراً، إذا كانت هناك درجتان متعادلتان، فإنه ينبغي علي الباحث أن يقوم بحساب متوسط الرتب المتعادلة، واعتماد قيمة المتوسط كرتبة نهائية. ومجموع الدرجات التالية تبين هذا الإجراء لعدد ثمان درجات N = 8:

الدرجات الأصلية	3	4	4	7	9	9	9	12
الوضع الترتيبي	1	2	3	4	5	6	7	8
الرتب النهائية	1	2,5	2,5	4	6	6	6	8

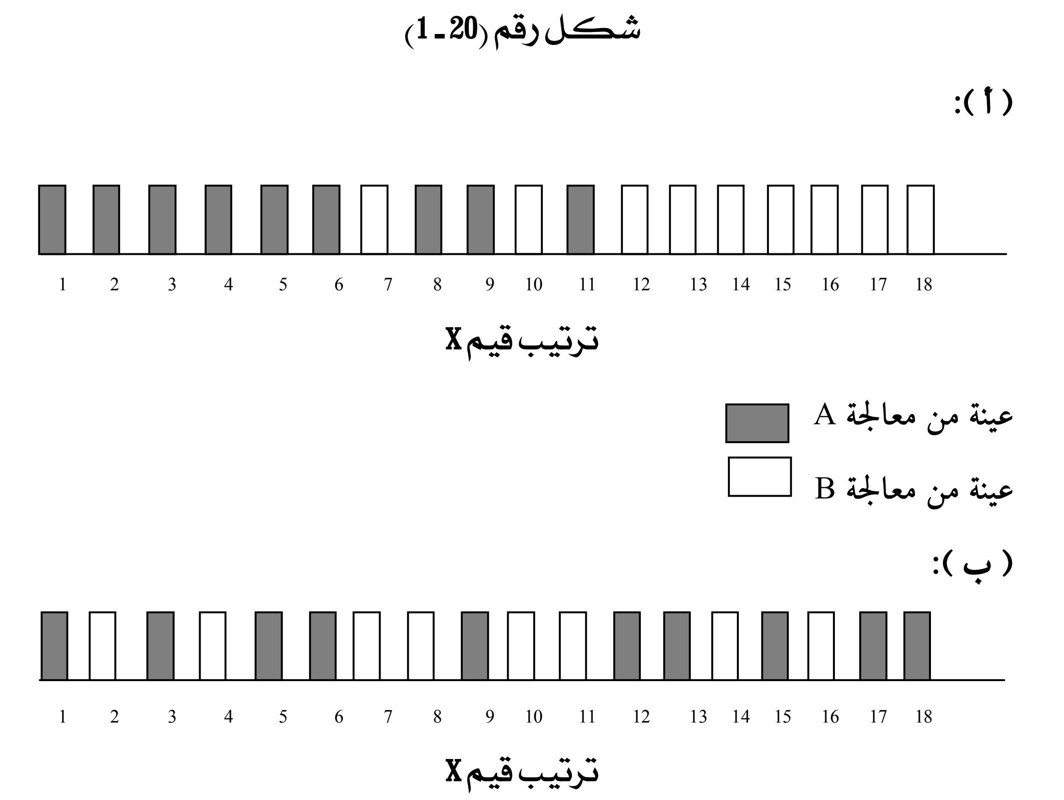
علي السبيل المثال، تحتوي الدرجات الأصلية علي شخصين درجاتهما متعادلة 4 = x وقد أعطي لهاتين الدرجتين رتب 2، 3، وبعد ذلك فإن كلا الشخصين قد أسند إليهما رتبة نهائية مساوية لمتوسط هذه الرتب (2.5=2 / 5=3+2). وهكذا في باقي الدرجات المتعادلة.

أولاً: اختبار مان- وتني The Mann-Whitney U كبديل لاختبار مقاييس t :

لقد صُمِّمَ هذا الاختبار لاستخدام بيانات من عينتين منفصلتين وذلك لتقييم الفرق بين مجتمعين لديها بيانات عددية، وغالباً ما يستخدم هذا الاختبار كبديل للاختبار t. ويتطلب حساب هذا الاختبار أن الدرجات الفردية في العينتين ينبغي أن ترتب، وأن العملية الرياضية لاختبار مان- وتني U تستند علي الملاحظة البسيطة التالية:

إن الفرق الحقيقي بين المجتمعين يجب أن يكون سبباً في وجود الدرجات في واحد من العينات تكون بشكل عام أكبر من الدرجات في العينة الأخرى. وإذا ما تم دمج عينتين لتصبحا عينة واحدة، ويتم ترتيب كل هذه الدرجات في صف واحد. وعندئذٍ فإن الدرجات من عينة واحدة يجب أن تتركز علي الطرف الأخر من الصف. وعلى الجانب الآخر، فإذا لم يكن هناك فرق، فإن الدرجات الكبيرة والصغيرة سوف تدمج بشكل متساوٍ في العينتين، لأنه لا يوجد أي سبب لأي واحد من مجموع الدرجات لأن يكون بشكلٍ منظم أكبر أو أصغر من مجموع الدرجات الأخرى. إن هذه الملاحظة يمكن التدليل عليها من خلال الشكل التالي:

شكل رقم (20 ـ 1)



المصدر: frederick J. GRavetter and larry B. wallnau , statistics for the Behavioral sciences , .8et ed, Wadsworth cengage learning , USA , 2010 , p.688

الفرضية الصفرية المرتبطة باختبار مان-وتني U :

لما كان اختبار مان- وتني U يقارن بين توزيعين (بدلاً من متوسطين)، فإن الفروض المتعلقة بهذا الاختبار تفضي إلى بعض الغموض، ويصاغ الفرض الصفري في إطار فرق متين ومنظم بين المجموعتين اللتَيْن يسعى الباحث للمقارنة بينهما.

الفرض الصفري H_O: لا يوجد فرق بين المعالجتين وعليه، لا يوجد أي غرض من وراء رتب حالة معالجة بأن تكون بشكل منظم أعلي (أو أقل) من رتب حالة المعالجة الأخرى.

الفرض البديل H_i: يوجد فرق بين المعالجتين. وعليه فإن رتب حالة إحدى المعالجات تكون بشكل منظم أعلى (أو أقل) من رتب حالة المعالجة الأخرى.

حساب قيمة اختبارات U:

إن الخطوة الأولى لحساب اختبار مان وتني U:

1- الحصول على عينتين منفصلتين. ونشير إلى عدد الأفراد في العينة الأولي بالرمز N_1،وعدد الأفراد في العينة التالية بالرمز N_2.

2- يتم دمج العينتين معاً N_1+N_2 وبعد ذلك يتم ترتيبهما.

مثال: إذا كان لدينا عينتان منفصلتان $N_1=6$ و $N_2=6$

N_1 = 15، 6، 48، 9، 2، 27

N_2 = 8، 94، 68، 18، 63، 71

أولا: نقوم باستخراج قيمة Uبحساب عدد الدرجات في العينة الأولى التي تسبق الدرجات في العينة الثانية، أي كم درجة في العينة الأولى يكون ترتيبها أقل من أي درجة من درجات العينة الثانية. ومن خلال ترتيب الدرجات تَبَيَّنَ لنا أن الدرجة الأولى في المجموعة الأولى لا يسبقها أية درجة من المجموعة الثانية أي أصغر وهكذا يمكننا الحصول على قيمة U_1 .

$$U_1 = 0 + 0 + 1 + 1 + 2 + 2$$
$$= 6$$

وبنفس الطريقة يمكننا حساب درجات العينة الثانية فنحصل كل قيمة U_2.

$$U_2 = 2 + 4 + 6 + 6 + 6 + 6$$
$$= 30$$

وبذلك يصبح لدينا قيمتان لـِ: U_1 و U_2

المعادلة التالية لهذه البيانات $U_1 + U_2 = n_1n_2$

$$6 + 30 = (6)(6)$$

ويمكننا أيضاً الحصول على قيمة U من نفس البيانات السابقة بإتباع الأتي:

نقوم أولاً بتنظيم هذه الدرجات الخاصة بهاتين العينتين وترتيبهما من أصغر درجة إلى أكبر درجة، ويتم ذلك بعد دمج درجات العينتين معاً.

أما حساب U من العينات الكبيرة، فإنه لما كانت عملية عد النقاط لكي نحدد قيمة مان - وتني هي عملية مضنية لا سيما عندما يكون حجم العينة كبيراً، فإننا نورد المعادلة التالية لتوليد قيمة U لكل عينة، ولاستخدام هذه المعادلة يتطلب الأمر دمج العينتين معاً، ومن ثم ترتيب كل الدرجات. وبعد ذلك ينبغي الحصول علي $\sum R_1$ أي مجموع ترتيب الأفراد في العينة N_1 ويقابلها $\sum R_2$ للعينة N_2 . وعندئذٍ يمكننا حساب قيمة U وفقاً للمعادلة التالية لكلتا العينتين.

$$U_1 = n_1(n_2) + \frac{n_1(n_1+1)}{2} - \sum R_1$$

$$U_2 = n_1(n_2) + \frac{n_2(n_2+1)}{2} - \sum R_2$$

من خلال البيانات في المثال السابق:

الدرجات	2	6	8	9	15	18	27	48	63	68	71	94
العينة	1	1	2	1	1	2	1	1	2	2	2	2
الترتيب	1	2	3	4	5	6	7	8	9	10	11	12

إن مجموع ترتيب الأفراد في العينة N_1 يكون:

$$\sum R_1 = 1 + 2 + 4 + 5 + 7 + 8 = 27$$

ومجموع ترتيب الأفراد في العينة N_2:

$$\sum R_2 = 3 + 6 + 9 + 10 + 11 + 12 = 51$$

وباستخدام المعادلة الخاصة بالعينة N_1 :

$$U_1 = n_1(n_2) + \frac{n_1(n_1 +_1)}{2} - \sum R_1$$
$$= 6(6) + \frac{6(7)}{2} - 27$$
$$= 36 + 21 - 27$$
$$= 30$$

وللعينة الثانية N_2:

$$U_2 = n_1(n_2) + \frac{n_2(n_2 +_1)}{2} - \sum R_2$$
$$= 6(6) + \frac{6(7)}{2} - 51$$
$$= 36 + 21 - 51$$
$$= 6$$

لاحظ أن هذه القيم هي نفس القيم التي تحصلنا عليها باستخدام الطريقة الأولى.

الاختبارات الصفرية من خلال اختبار مان - وتني U:

تجدر الإشارة إلى أنه حتى الآن قد تم تطوير طريقة لبيان نظام الترتيب من خلال قيمة عددية. وعليه فإن المشكلة اللاحقة هي أن يقرر الباحث ما إذا كانت قيمة U تقدم لنا البرهان إن كان هناك فرق حقيقي بين المجموعتين. ولما كان الفرض الصفري لاختبار مان - وتني U يصاغ بأنه لا يوجد فرق نظامي بين العينتين اللتين تخضعان للمقارنة؛ أي لا يوجد فرق حقيقي بين المجتمعين اللَّذَيْن تم سحب العينة منهما. في الحالة الراهنة، فإن النتيجة الأكثر ترجيحاً أن العينتين تكونان متماثلتين، وأن قيمة U تكون كبيرة نسبياً. وعلى الجانب الآخر، أنه عندما تصل قيمة U قريبة من الصفر، يكون هناك نزعة نحو رفض الفرض الصفري.

إن التوزيع الاحتمالي لكل قيم U قد تم بناؤه. وأن القيم الحرجة لمستوى الدلالة $\alpha = .05$، ومستوى الدلالة $\alpha = .05$ يمكن الرجوع إليهما من خلال الجدول (7) انظر ملحق (1). وعندما نتحصل على قيمة U من بيانات العينة أصغر أو مساوية للقيمة الجدولية حينئذٍ يمكننا رفض الفرض الصفري (هذا عكس ما تعودنا عليه في الفرض الصفري عند التعامل مع الاختبارات)، فعلى سبيل المثال، في المثال الذي بين أيدينا نجد أن كلتا العينتين لديها: N تساوي 6 درجات، وأن القيمة الجدولية لـِ U تساوي 5 (اختبار ثنائي الجانب) بمستوي دلالة $\alpha = .05$. هذا يعني أن البيانات التي تحصلنا عليها لا تقدم لنا البرهان الكافي لكي نصل إلى نتيجة مفادها أن هناك فرقاً دالاً بين المجموعتين.

التقدير التقريبي الطبيعي لاختبار مان ـ وتني U:

عندما تكون لدينا عينتان كبيرتان N=20، ويكون الفرض الصفري فرضاً صحيحاً، فإن توزيع إحصاء U يميل إلى أن يكون التوزيع قريباً من الشكل الطبيعي approximate normal shape. في هذه الحالة فإنه باستطاعتنا تقييم الفروض المرتبطة باختبار مان- وتني من خلال استخدام إحصاء درجة Z-Score ، أي تحويل درجة U إلى درجة معيارية (Z)، ويمكننا هنا الإشارة إلى أن جدول القيم الحرجة لاختبار مان - وتني لا يتضمن قيماً لعينات أكبر من N=20، ويرجع السبب في ذلك إلى أن التقدير التقريبي الطبيعي نموذجيٌّ Typically يستخدم مع العينات الكبيرة.

إن الإجراء المتبع لهذا التقدير التقريبي الطبيعي هو كالتالي:

1- إيجاد قيم U لعينة N_1، والقيمة N_2 كما فعلنا سلفاً، وتعتبر قيمة U هي القيمة الأصغر من هاتين القيمتين.

2- عندما يكون حجم العينتين كبيراً نسبياً (عدد الأفراد 20 أو أكثر) فإن توزيع إحصائي اختبار مان - وتني U يميل ليشكل توزيعاً طبيعياً:

$$\sigma = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} \quad \text{و} \quad U = \frac{n_1 n_2}{2}$$

ويمكننا الحصول علي U من بيانات العينة التي يمكن تحديدها في هذا التوزيع مستخدمين درجة Z (Z-Score) :

$$Z = \frac{x - u}{\sigma} = \frac{\dfrac{U - n_1 n_2}{2}}{\sqrt{\dfrac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

3- وباستخدام وحدة الجدول الطبيعي لتعيين المنطقة الحرجة لدرجة Z، علي سبيل المثال، فإن القيمة الحرجة لـِ Z بمستوى دلالة 05. تكون 1.96 ±.

مثال تطبيقـي:

لتوضيح التقدير التقريبي الطبيعي، فإنه بإمكاننا التعامل مع نفس البيانات السابقة، حيث تهدف البيانات السابقة إلى مقارنة عينتين 1 و 2 مستخدمين عينتين مستقلتين يصل عدد كل منها إلي 6. وأن هذه البيانات قد ولدت قيمة تساوي ستة U=6 (لاحظ أنه عندما تكون العينات صغيرة، فإننا في العادة نستخدم جدول القيم الحرجة لاختبار مان- وتني بدلاً من جدول التقدير التقريبي الطبيعي، وعلى أية حال، فإننا نستخدم التقدير الطبيعي حتى نتمكن من مقارنة النتيجة المتحصل عليها بالنتيجة التي تحصلنا عليها من خلال اختبار مان- وتني الاعتيادي). وباستخدام المعادلة أعلاه، فإن درجة Z المناظرة لـِ: U=6 تكون:

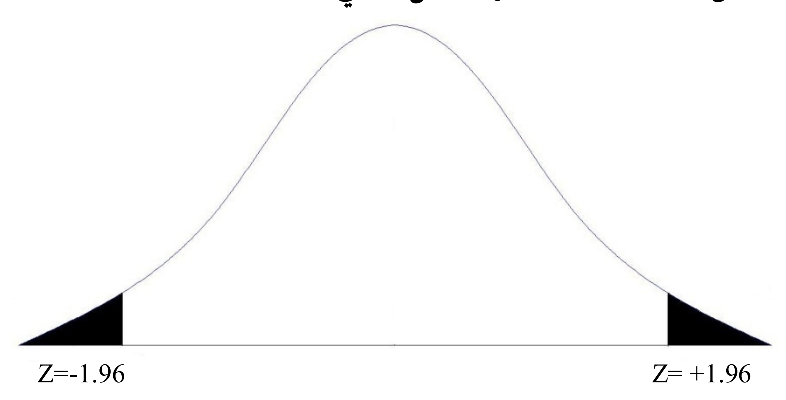
$$Z = \frac{u - n_1 n_2}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

$$= \frac{6 - (6)(6)}{\sqrt{\frac{(6((6)(6 + 6 + 1)}{12}}}$$

$$= \frac{-12}{\sqrt{\frac{468}{12}}}$$

$$= -1.92$$

وطبقاً لوحدة الجدول الطبيعي فإن 5 % المتطرفة للتوزيع الطبيعي تكون موجودة في الذيلين أبعد من 1.96 ±=Z (انظر الشكل التالي:)



بدرجة دلالة .50 = ∝، و أن درجة Z أبعد من 1.96 ±فهي بالتالي تقود إلى رفض الفرق الصفري. إن القيم المحسوبة لـِ Z=-1.92 لا تقع في المنطقة الحرجة، وعليه فإن القرار الذي يمكن أن نصل إليه مفاده أنه ليس بإمكاننا رفض الفرض الصفري. (لاحظ أننا قد توصلنا إلى نفس النتيجة للاختبار الأصلي باستخدام القيمة الحرجة لجدول اختبار مان – وتني U).

إطار:

إعداد تقرير حول نتائج اختبار U :

بخلاف كثير من النتائج الإحصائية الأخرى، لا توجد قواعد صارمة لإعداد تقرير حول نتائج اختبار مان – وتني U. وبناءً على دليل الجمعية الأمريكية لعلم النفس APA فإنه يقترح أنه ينبغي في إعداد التقرير أن يشمل ملخصاً للبيانات (مثل: معلومات حول حجم العينة، ومجموع الرتب)، والإحصاء المتحصل عليه، وقيمة P. ففي المثال الذي بين أيدينا، فإن نتائج الدراسة يمكن أن تعد في شكل التقرير التالي:

الدرجات الأصلية قد تم ترتيبها، وتم حساب اختبار U لمقارنة الرتب لعدد 6 مشاركين في المعالجة الأولى (أ)، و 6 مشاركين في المعالجة الثانية (ب). وقد أشارت النتيجة إلى أنه يوجد فرق دال بين المعالجتين $U=6, P > 0.05$ لمجموع رتب مساوية لـِ 27 للمعالجة (أ)، و 51 للمعالجة (ب) [2].

الافتراضات والمحاذير لاستخدام اختبار مان ـ وتني U:

يعتبر اختبار مان – وتني U اختباراً مفيداً جداً كبديل لقياس اختبار T المستقل، باعتبار أن اختبار U لا يتطلب تجانسية التباين Homogeneity of Variance أو التوزيعات الطبيعية، ومن هنا يمكن استخدامه في المواقف التي لا يكون فيها استخدام الاختبار ممكناً، وعلى أية حال، فإن اختبار U لا يتطلب مشاهدات مستقلة Independent observations، وإنما يفترض أن يكون المتغير التابع متغيراً متصلاً (المتغير

المتصل هو ذلك المقياس الذي يمتلك عدداً لا متناهياً من النقاط الواضحة المعالم). إن أحد نتائج هذه الحقيقة أنه من غير المرجح لشخصين أن يمتلكا بالضبط نفس الدرجة. وهذا يعني أنه يجب أن يوجد قليل من الدرجات المتعادلة في البيانات. وعندما تكون لدينا بيانات عينة لديها عدة درجات متعادلة، فهذا يعني أن الافتراض الأساسي لاختبار مان- وتني U قد انتهك. وفي هذه الحالة ينبغي على الباحث أن يكون حذراً في استخدامه لاختبار مان - وتني U. وتجدر الإشارة هنا إلى أنه إذا كان هناك قليل من الدرجات المتعادلة نسبياً يمكن للباحث استخدام اختبار U، لكن ينبغي عليه اتباع الإجراء المعياري لترتيب الدرجة المتعادلة [4].

ثانياً: اختبار الإشارة والرتب ولكوكسن (T) Wilcoxon Test:

لقد صُمِّمَ هذا الاختبار لتقييم الفرق بين معالجتين باستخدام بيانات جاءت من تجربة متكررة القياس a repeated - measures. لاحظ أن دراسة القياسات المتكررة يتعلق فقط بعينة واحدة حيث إن كل فرد في العينة يتم قياسه مرتين: مرة في المعالجة الأولى، والمرة الأخرى في المعالجة الثانية. والفرق بين القياسين يتم تسجيله كدرجة تتعلق بذلك الفرد.

ويتطلب اختبار ولكوكسن ترتيب الفرق بشكل منظم من الأصغر إلى الأكبر From Smalest to Largest بغض النظر عن الإشارة أو الاتجاه. علي سبيل المثال، في الجدول الموالي (أ) يبين لنا الفروق غير العددية Nonnumerical differences والرتب Ranks لعينة تحتوي على أربعة مشاركين N=4. فإذا كانت الفروق تمثل قيماً عددية Numerical Values، فيمكن أيضاً ترتيبها في إطار حجمها المطلق absolute magnitude. أما الجدول (ب) فيوضح لنا أن N=5 مع فرق الدرجات العددية وترتيبها.

جدول (أ) فرق الدرجات في الجزء (أ) الفروق يتم قياسها كزيادة نسبية أو نقصان، والقيم غير العددية تم ترتيبها طبقاً للحجم، مستقلة في الاتجاه.

أما الجزء (ب) فإن فرق الدرجات يمثل قيماً عددية، والفرق تم ترتيبه مستقلاً في الاتجاه.

(أ)

المشاركون	الفرق من المعالجة الأولى للمعالجة الثانية	الرتبة
أ	زيادة صغيرة	2
ب	نقص كبير جداً	4
ج	زيادة متوسطة	3
د	زيادة صغيرة جداً	1

(ب)

المشاركون	فرق الدرجة	الرتبة	
أ	+ 4	2	$+\sum R = 6$
ب	- 14	5	
ج	+ 9	4	$-\sum R = 9$
د	- 1	1	
هـ	- 6	3	T = 6

المصدر: frederick J.Gravetter and Larry B.wallnau statistics for the Behavioral sciences , 8th ed ,wadsworth cengage learning,USA ,2010,p 675.

الفروض المتعلقة باختبار ولكوكسن :T

يصاغ الفرض الصفري لاختبار ولكوكسن ببساطة، أنه لا يوجد فرق ثابت ونظامي بين المعالجتين.

H_O: لا يوجد فرق بين المعالجتين. وعليه، فإنه في المجتمع العام لا يوجد اتجاه في الفروق في الدرجات إما أن تكون بشكل نظامي موجبة أو بشكل نظامي سالبة.

H_i: يوجد فرق بين المعالجتين. وعليه، فإنه في المجتمع العام يكون اتجاه الفروق في الدرجات إما بشكل نظامي موجب أو بشكل نظامي سالب.

فإذا كان الفرض الصفري (H_O) فرضاً صحيحاً، فإن أيّاً من الفروق التي توجد في بيانات العينة يجب أن يكون مردها لعامل الصدفة، وعليه، فإننا بذلك سنتوقع فروقاً موجبةً وفروقاً سالبةً تكون متمازجة بالتساوي. intermixed evenly. وعلى الجانب الآخر، إذا كان الفرق في الدرجات فرقاً ثابتاً ونظامياً بين المعالجتين، فإنه بالتالي يجب أن يتسبب في وجود درجات كبيرة في إحدى هذه المعالجات مقارنة بالدرجات في المعالجة الأخرى.

وهذا يجب أن يقود إلى توليد فروق في الدرجات التي تتجه بشكل ثابت نحو الموجب أو السالب. ويستخدم اختبار ولكوكسن T الإشارات (S) والرتب RANKS في الفرق في الدرجات لتقييم ما إذا كان هناك فرقُ ذو دلالة إحصائية بين المعالجتين(6).

حساب وتفسير اختبار ولكوكسن (T):

كما هو الحال في معظم الاختبارات غير البارامترية، فإن حساب اختبار ولكوكسن يكون إلى حد بعيد بسيطاً في عملية حسابه حيث:

1- ترتيب القيم المطلقة لفرق الدرجات.

2- تقسيم الترتيب إلى مجموعتين: مجموعة مرتبطة بالفروق الموجبة (الزيادة) والأخرى مرتبطة بالفروق السالبة (النقصان).

3- إيجاد مجموعة الرتب لكل مجموعة، واعتماد المجموع الأصغر لهاتين المجموعتين كإحصائي الاختبار لاختبار ولكوكسن. ويشار إلى ذلك بالرمز T.

ففي البيانات في الجدول السابق (أ)، على سبيل المثال، فإن الزيادة في الترتيب كانت متعلقة بـِ: 1، 2، و 3 التي يصل مجموعتها إلى $\sum R = 6$. في حين يوجد نقص واحد فقط المرتبط بالترتيب 4. ومن هنا فإن قيمة T لمجموعة هذه البيانات: T=4. أما في الجدول (ب)، فإن الفرق في الدرجات (الدرجات الموجبة) فهي مرتبطة بالترتيب 2، و4 التي يصل مجموعها إلى: $\sum R = 6$. وأن فرق الدرجات التي تحمل إشارة سالبة فهي مرتبطة بـِ: 1،3، و5 التي يصل مجموعها إلى $\sum R = 9$، وأن قيمة T لهذه الدرجات وصلت إلى: T= 6

وكما بينا سابقاً، فإن أقوى تأثير معالجة يجب أن يسبب الفرق في الدرجات بحيث يكون فرقاً موجباً أو فرقاً سالباً. وأنه في الحالة المتطرفة فإن كل الفروق تكون في نفس الاتجاه مولدة T=0. فعلى سبيل المثال، عندما تكون كل الفروق فروقاً موجبة، فإن مجموع الرتب السالبة تكون صفراً. وعلى الجانب الآخر، إذا لم يكن هناك أي تأثير للمعالجة N0 treatment effect. فإن الإشارات المتعلقة بالفرق في الدرجات يجب أن تتمازج بالتساوي. وفي هذه الحالة يكون اختبار ولكوكسن T كبيراً جداً نسبياً. وبشكل عام، فإن قيمة T الصغيرة (قريبة من الصفر) تقدم دليلاً للفرق الحقيقي بين حالتي المعالجة. إن توزيع كل القيم المحتملة لِـ: T قد تم بناؤه. وأن القيم الحرجة لمستوي دلالة (α) 0.05 ومستوى دلالة 0.01 يمكن الحصول عليها من الجدول المتعلق بقيم T الحرجة (انظر الملحق رقم 1) وكلما ولدت بيانات عينة أقل أو مساوية للقيمة الحرجة، فإن ذلك يقود إلى رفض الفرض الصفري H_O . ولمقاييس الدراسة المتكررة لعينة تصل إلى N=15.15، علي سبيل المثال، فإن الجدول المتعلق باختبار ولكوكسن للقيم الحرجة T=25 اختبار ثنائي الجانب على مستوى دلالة .05 = α. وهذا يعني أن قيمة T الحرجة التي تساوي 25 أو أصغر تكون أكثر ترجيحاً أن تحدث (احتمالية .05 أو أقل) إذا كان الفرض الصفري صحيحاً[7].

الدرجات المتعادلة ودرجات صفر:

بالرغم من أن اختبار ولكوكسن T لا يتطلب حسابه التوزيعات الطبيعية Normal distributions، إلا أنه يفترض أن يكون المتغير التابع متغيراً متصلاً. وكما نوهنا سلفاً عند الحديث عن اختبار مان- وتنى، فإن هذا الافتراض ينطوي علي أن تلك الدرجات المتعادلة ينبغي أن تكون إلى حد كبير بعيدة الاحتمال. فعندما يظهر التعادل في بيانات العينة، ينبغي على الباحث أن يكون قلقاً بأن فرضية المتصلية Continuity قد تم انتهاكها. ومن هنا قد لا يكون اختبار ولكوكسن T اختباراً مناسباً.

أما إذا كان هناك تعادل قليلُ نسبياً في البيانات، فإن معظم الباحثين يفترضون أن البيانات هي حقاً بيانات متصلة ولكن قد تم قياسها بشكل فج، في هذه الحالة، يمكن

للباحث استخدام اختبار ولكوكسن T و إبلاء القيم المتعادلة انتباهاً خاصاً عند العمليات الحسابية[8].

تجدر الإشارة إلى أن هناك نمطين مختلفين للدرجات المتعادلة:

1- إن الموضوع الخاضع للدراسة قد تكون درجته متساوية في المعالجة الأولى وفي المعالجة الثانية، وينتج عن ذلك فرق يساوي صفراً (0).

2- قد يكون لدينا موضعان أو أكثر متماثلان في فرق الدرجات (متجاهلين علامة هذا الفرق)[9].

وعندما تحتوي البيانات أفراداً بفروق درجات تساوي صفراً (0)، حينئذٍ ينبغي طرح أولئك الأفراد جانباً من التحليل، وخفض حجم العينة المدروسة (N) ومع ذلك، فإن هذا الإجراء يتجاهل حقيقة أن فرق الدرجة التي تساوي صفراً (0) يكون دليلاً يبقى على الفرضية الصفرية (Ho). إن أفضل إجراء يمكن اتباعه هو تقسيم فروق الدرجات التي تساوي صفراً بشكل متساوٍ بين الدرجات الموجبة والدرجات السالبة (إذا كان لدينا رقم مفرد لفروق الدرجات التي تساوي صفراً 0). فإنه ينبغي علينا أن نتجاهل أحدهما ويقسم الباقي بشكل متساوٍ) إن هذا الإجراء الثاني يهدف إلى زيادة $\sum R$ لكل من الرتب الموجبة والرتب السالبة التي تزيد من القيمة النهائية (T)، وتجعل هذه القيمة أكثر ترجيحاً على أنه سوف نبقي على الفرض الصفري.

وعندما تكون لدينا درجات متعادلة بين فروق الدرجات، فإن كل واحدة من الدرجات المتعادلة ينبغي أن يخصص لها الوسط الحسابي للرتب المتعادلة.

اختبار الفرض باستخدام ولكوكسن T:

المثال الموالي يعطينا مثالاً بيّناً لاختبار ولكوسن باستخدام بيانات لنمطين من درجات متعادلة.

مثال: البيانات التالية تمثل عشر هيئات. وكل هيئة من هذه الهيئات كانت تهدف إلي زيادة نسبة العاملين لديها الذين يشاركون في برنامج للتبرع بالدم. البيانات التالية توضح

نسبة المشاركين في البرنامج في السنة الماضية (قبل الشروع في الدعاية لهذا البرنامج). وخلال هذا العام وباستخدام اختبار ولكوكسن T لتقييم ما إذا كانت البيانات تقدم دليلاً كافياً فإن الدعاية لهذا البرنامج قد أحدثت تأثيراً دالاً على عملية التبرع بالدم. لقد رُتِّبَتْ بيانات الجدول بشكل منظم وذلك طبقاً للقيمة المطلقة لفرق الدرجات.

نسبة المشاركين					
الهيئة	قبل الدعاية	بعد الدعاية	الفرق	التركيب تجاهل الأصفار	التركيب تضمين الأصفار
1	18	18	0	–	1.5
2	24	24	0	–	1.5
3	31	30	−1	1	3
4	28	24	−4	2	4
5	17	24	+7	3	5
6	16	24	+8	4	6
7	15	26	+11	5.5	7.5
8	18	29	+11	5.5	7.5
9	20	36	+16	7	9
10	9	28	+19	8	10

المصدر: Ibid , P. 677.

أـ تجاهل الفروق التي تساوي صفرا (0):

إن أولى الخطوات لإجراء اختبار ولكوكسن T هي الأخذ بالتوصية المتعلقة بتجاهل الفروق التي تساوي صفراً.

الخطوة الأولى: صياغة الفرض الصفري:

H_O: لا يوجد تأثير للدعاية لهذا البرنامج، عليه، فإن أية فروق في الدرجات يكون مرده لعامل الصدفة. وهنا ينبغي ألا يكون هناك نمطٌ ثابت لهذه الفروق.

الخطوة الثانية: إن الهيئتين اللَّتَيْن لديهما فرق في الدرجات مساوياً لصفر ينبغي تجاهلها. وأن حجم العينة بالتالي ينخفض إلى: N=8 ودرجة الدلالة 05. = α. والقيمة الحرجة لاختبار ولكوكسن T يساوي 3.

إن أي قيمة لإحصائي الاختبار أقل أو مساوية لـ: 3 تقود إلى رفض الفرض الصفري.

الخطوة الثالثة: تحدد مجموع الفروق ذات الرتب الموجبة (+):

$$3 ، 4 ، 5 ، 5 ، 5 ، 5 ، 7 ، 8 . \sum R+ = 33$$

ومجموع الفروق ذات الرتب السالبة (−):

$$1 ، 2 ، . \sum R - = 3$$

بعد ذلك يتم اعتماد أصغرهما. عليه: T=3

الخطوة الرابعة: لما كانت قيمة T=3 أصغر من القيمة الحرجة لـ:T؛ عليه نرفض الفرق الصفري ونخلص إلى القول بأنه يوجد تغيير دال في المشاركة بالتبرع بالدم في حملة الدعاية التي قام بها الهلال الأحمر [10].

ب- تضمين الفروق التي تساوي صفرا (0):

إذا أضفنا الشخصين اللَّذَيْن لديهما فرق الدرجات المساوية لصفر، حينئذٍ يصبح حجم العينة N=10 بدرجة دلالة 0.55 = α والقيمة الحرجة لاختبار ولكوكسن T هي 8. ومهما كانت فروق الدرجات التي تساوي صفراً متعادلةً من حيث الترتيب الأول والثاني، فقط أَعْطِ لكل منهما 1.5 وأن أحد هذه الرتب تم تخصيصها للمجموعة الموجبة، و الأخرى للمجموعة السالبة كما هو مبين في النتيجة التالية:

$$\sum R+ = 1.5 + 5 + 6 + 7.5 + 7.5 + 9 + 10$$

$$= 46.5$$

و $$\sum R- = 1.5 + 3 + 4$$

$$= 8.5$$

ولما كانت قيمة ولكوكسن T أصغر مجموع لهاتين القيمتين T=8.5 ولما كانت هذه القيمة أكبر من القيمة الحرجة، عليه نقبل الفرض الصفري.

ونَخْلُصُ إلي القول بأن هذه البيانات لم تقدم لنا دليلا كافياً على أن هناك تغيراً دالاً لدى المشاركين في برنامج التبرع بالدم.

تجدر الإشارات هنا إلي أنه بإضافة فروق الدرجات التي تساوي صفراً (0) للاختبار قد غيرت من النتيجة الإحصائية. فالفروق في الدرجات التي تساوي صفراً (0) هي في الواقع تؤشر بأن الفرض الصفري فرضُ صحيحُ.

التقدير التقريبي الطبيعي لاختبار ولكوكسن (T): [11]

عندما تكون العينة كبيرة نسبياً، فإن قيم إحصاء ولكوكسن T تميل إلي تشكيل توزيع طبيعي.في هذه الحالة، فإن من الممكن إجراء الاختبار وذلك باستخدام إحصاء درجة Z، والتوزيع الطبيعي،بدلاً من النظر إلي قيمة T في جدول ولكوكسن. وعندما يكون حجم العينة أكبر من 20، فإن التقدير التقريبي الطبيعي يكون دقيقاً وأنه بالإمكان استخدامه.ففي العينات الكبيرة التي تتعدي N=>50، فإن الجدول المتعلق بدرجات ولكوكسن لا يمدنا علي نحو نموذجي بأية قيم حرجة.عليه، يصبح من الضرورة بمكان استخدام التقدير التقريبي الطبيعي.

إن الإجراء المتبع للتقدير التقريبي لاختبار ولكوكسن (T) هو كالتالي:

1- إيجاد مجموع الرتب ذات الإشارات الموجبة (+) وإجمالي الرتب التي تحمل الإشارات السالبة (-) كما بينا سابقاً. وأن قيمة ولكوكسن (T) تمثل أصغر القيمتين.

2- عندما تكون (N) أكبر من 20،فإن قيم ولكوكسن (T) تشكل توزيعاً طبيعياً بمتوسط:

$$U = \frac{n(n+1)}{4}$$

وانحراف معياري: $$\sigma = \sqrt{\frac{n(n+1)(2N+1)}{24}}$$

في هذه الحالة يتم تحويل ولكوكسن (T) إلي درجة Z :

$$Z = \frac{x-u}{\sigma} = \frac{\dfrac{T - N(n+1)}{4}}{\sqrt{\dfrac{n(n+1)(2N+1)}{24}}}$$

3- يستخدم جدول المنحي الطبيعي المعياري لتحديد المنطقة الحرجة لدرجة Z. علي سبيل المثال، القيم الحرجة تكون 1.96± بدرجة دلالة 05. = α.

مثال تطبيقـي:

كما أشرنا في موضعه إلى أن التقدير التقريبي لاختبار (T) لعينة أكبر من 20 فرداً، أنه سوف تبرهن لنا العمليات الحسابية لنفس بيانات المثال المنفرط الذي قدمناه كمثال لحساب (T).

أـ تجاهل الفروق التي تساوي صفرا (0):

عندما تم تجاهل الفروق التي تساوي صفراً (0) لتلك الهيئتين، قد تحصلنا علي N=8 و T=3. وباستخدام التقدير التقريبي الطبيعي فإن تلك القيم تولد:

$$u = \frac{(n+1)}{4} = \frac{8(9)}{4} = 18$$

$$\sigma = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

$$= \sqrt{\frac{8(9)(17)}{24}} = \sqrt{51} = 7.14$$

وبهذه القيم فإن T التي تحصلنا عليها تقابل درجة Z التالية:

$$Z = \frac{T-u}{\sigma} = \frac{3-18}{7.4} = \frac{-15}{7.14} = -2.10$$

ومع الحدود الحرجة 1.96 ±، فإن درجة Z المتحصل عليها تكون قريبة من الحد، ولكنها كافية لأن تكون دالة على مستوى 05. = α. لاحظ أن هذه النتيجة مساوية

بالضبط لتلك النتيجة التي توصلنا إليها باستخدام جدول توزيع ولكوكسن T عندما تجاهلنا الفروق التي تساوي صفراً (0) في المثال المنفرط.

تضمين الفروق التي تساوي صفرا (0):

إذا أضفنا الشخصين اللّذَيْن لديهما فرق الدرجات المساوية لصفر (0)؛ حينئذٍ يصبح حجم العينة N = 10 و T = 8.5. وباستخدام التقدير التقريبي الطبيعي، فإن هذه القيم تولد:

$$u = \frac{(n+1)}{4} = \frac{10(11)}{4} = 27.5$$

$$\sigma = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

$$= \sqrt{\frac{10(11)(21)}{24}} = \sqrt{96.25} = 9.81$$

وبهذه القيم فإن قيمة T المتحصلين عليها تماثل درجة Z التالية:

$$Z = \frac{x-u}{\sigma} = \frac{8.5-27.5}{8.81} = \frac{-19}{9.81} = -1.94$$

ومع الحدود الحرجة 1.96 ±، فإن درجة Z التي تحصلنا عليها تكون قريبة من الحد، ولكنها ليست كافية لأن تكون دالة على مستوى $\alpha = .05$. مرة ثانية، إن هذه النتيجة هي تلك النتيجة نفسها التي توصلنا إليها مستخدمين جدول T، عند تضميننا للفروق التي تساوي صفراً (0) في المثال السابق.

ولإعداد تقرير حول النتائج المتوصل إليها من خلال استخدام هذا الاختبار (T)، يمكننا القول بإنه لا يوجد شكل محدد لإعداد تقرير حول نتائج اختبار ولكوكسن (T)، فإنه يقترح أن يشتمل هذا التقرير على ملخص للبيانات والقيمة المتحصل عليها من إحصاء هذا الاختبار. إضافة إلى قيمة P. كذلك يقترح أن يصف التقرير المعالجة التي تمت بخصوص فروق الدرجات التي تساوي صفراً (0) في التحليل. ففي الدراسة التي أوردناها في هذا الجزء من هذا الفصل، بينت أن هناك هيئتين لا تَغَيُّرَ في مشاركتهما في حملة التبرع

بالدم؛ وبالتالي قد تم تجاهلهما قبل عملية التحليل. وبالتالي تم ترتيب ثمان مجموعات فقط (N=8) وفقاً للمتغير في مستوى المشاركة. وقد تم استخدام اختبار T لتقييم هذه البيانات. وقد بَيَّنَتْ النتائج أن هناك زيادةً دالةً في المشاركة بعد الحملة الدعائية: $T = 3, p < .05$ للرتب بزيادة تصل إلى 33، وللرتب بالنقصان تصل إلى 3.

ثالثاً: اختبار كروسكال وليز Kruskal - Wallis Test:

يستخدم اختبار كروسكال - وليز لتقييم الفروق بين ثلاث عينات أو أكثر مستخدماً بيانات جاءت من خلال تصميم المقاييس المستقلة. ويعتبر هذا المقياس مقياساً بديلاً لتحليل التباين أحادي الجانب ANOVA.

وإذا كان مقياس أنوفا يتطلب إجراؤه درجات عددية يمكن استخدامها لحساب المتوسطات والتباين؛ فإن اختبار كروسكال - وليز على الجانب الآخر، يتطلب، ببساطة، ترتيب الأفراد بشكل منظم ترتيباً تصاعدياً من أصغر رتبة حتى أكبرها على المتغير المطلوب قياسه. وتجدر الإشارة هنا إلى أن اختبار كروسكال - وليز شبيه لاختبار مان - وتني الذي تناولناه في بداية هذا الفصل. إلا أن الأخير يستخدم للمقارنة بين مجموعتين، في حين أن اختبار كروسكال - وليز يستخدم في تقييم الفروق بين ثلاث مجموعات أو أكثر. إضافة إلى أن مقياس كروسكال - وليز يمكن استخدامه إذا كانت البيانات الأصلية العددية قد تحولت إلى قيم ترتيبية. المثال التالي يوضح العملية التي من خلالها يتم تحويل الدرجات العددية إلى رتب يمكن استخدامها في تحليل كروسكال - وليز.

مثال: يوضح المثال التالي البيانات الأصلية لدراسة تتعلق بثلاث مجموعات. ولأجل إعداد البيانات بشكل يمكن استخدامه في اختبار كروسكال - وليز، يتطلب بادئ ذي بدء ترتيب كل الدرجات الأصلية وفقاً للطريقة المعتادة لترتيب الدرجات المتعادلة؛ أي أن كل الدرجات الأصلية بعد ذلك تستبدل برتبها.

جدول رقم (20ـ 2)

البيانات المعدة للتحليل باستخدام اختبار كروسكال ـ وليز. البيانات الأصلية تحتوي على درجات عددية كما هو مبين في الجدول (أ)، والبيانات الأصلية سجلت بشكل منظم وخصصت لها رتب، والرتب أصبحت بديلاً للدرجات الأصلية كما هو مبين في الجدول (ب)

رتب البيانات الترتيبية (ب)

3	2	1
15	1	9
5	9	2
9	6	14
12	7	3.5
13	3.5	11

إيجاد ترتيب كل درجة

الترتيب الأصلي	الدرجات الأصلية العددية
1	2
2	3
3.5	5
3.5	5
5	8
6	9
7	12
9	14
9	14
9	14
11	16
12	19
13	20
14	21
15	26

البيانات الأصلية للدرجات العددية (أ)

3	3	1
26	2	14
8	14	3
14	9	21
19	12	5
20	5	16

المصدر: Frederick J. Gravetter and Larry B. Wallnau , op.cit , P. 682.

الفرضية الصفرية لاختبار كروسكال - وليز:

كما هو الحال في كل الاختبارات المتعلقة بالبيانات الترتيبية، فإن الفرضية لاختبار كروسكال - وليز تميل إلى شيء من الغموض. وبشكل عام، يصاغ الفرض الصفري لاختبار كروسكال - وليز بأنه لا توجد فروق ذات دلالة إحصائية بين المجموعات الخاضعة للمقارنة. وبشكل أكثر تحديداً، يصاغ الفرض الصفري (Ho) أنه لا يوجد ميل بأن رتب أي مجموعة ستكون بشكل منظم أعلى (أو أقل) من رتب أي مجموعة أخرى. عليه، يمكن صياغة الفرض الصفري والفرض البديل بالشكل التالي:

Ho: لا يوجد ميل للرتب في أي مجموعة أن تكون بشكل منظم أعلى أو أقل من رتب أية مجموعة أخرى.

Hi: أن الرتب على الأقل في واحدة من المجموعات تكون بشكل منظم أعلى أو أقل من رتب في مجموعة أخرى. وبالتالي توجد فروق بين المجموعات.

المعادلات المرتبطة باختبار كروسكال - وليز:

الجدول التالي يبين لنا تحويل البيانات من بيانات عددية إلى رتب. ويتطلب الأمر عند تطبيق المعادلات المتعلقة بهذا الاختبار اتباع الإجراءات التالية:

1- تضاف الرتب إلى بعضها البعض لكل مجموعة للحصول على العدد الإجمالي أو قيمة T لحالة المعالجة. وأن هذه القيمة المتحصل عليها (T) يتم استخدامها في معادلة كروسكال - وليز.

2- عدد الحالات لكل مجموعة يشار إليها بـِ (n) الصغيرة.

3- العدد الإجمالي للمجموعات بكاملها يشار إليه بـِ (N) الكبيرة.

وتولد لنا معادلة كروسكال - وليز إحصاءً يشار إليه بـِ (H) ولديه تقريباً نفس توزيع مربع كاي (X^2) بدرجة حرية تعرف بـِ: عدد المجموعات ناقص واحد.

الجدول (20ـ3) التالي:

1	2	3
9	1	15
2	9	5
14	6	9
3.5	7	12
11	3.5	13
$T_1 = 39.5$	$T_2 = 26.5$	$T_3 = 54$
$n_1 = 5$	$n_2 = 5$	$n_3 = 5$

وللبيانات الواردة في الجدول أعلاه، فإن عدد الحالات 3، وبالتالي فإن معادلة مربع كاي تولد لنا قيمة متعلقة بمربع كاي بدرجة حرية 2.

والمعادلة المتعلقة بإحصاء كروسكال - وليز هي كالتالي:

$$H = \frac{12}{N(N+1)}\left(\sum \frac{T^2}{n}\right) - 3(N+1)$$

وباستخدام البيانات السابقة، فإن معادلة كروسكال - وليز تولد لنا قيمة مربع كاي:

$$H = \frac{12}{15(16)}\left(\frac{(39.5)^2}{5} + \frac{(26.5)^2}{5} + \frac{(54)^2}{5}\right) - 3(16)$$

$$= 0.05\ (312.05 + 140.45 + 583.2) - 48$$

$$= 51.785 - 48$$

$$= 3.785$$

وبدرجة حرية (2) df=2، فإن القيمة الحرجة لمربع كاي تساوي 5.99 بدرجة دلالة $\alpha = .05$. ولما كانت القيمة المحسوبة لمربع كاي (3.785) أصغر من القيمة الجدولية؛ فإننا بالتالي نكون غير قادرين على رفض الفرض الصفري (Ho) حيث إن البيانات لم تقدم لنا دليلاً كافياً يمكننا من القول بأنه توجد فروق دالة بين هذه المجموعات الثلاثة.

ولإعداد تقرير حول نتائج اختبار (H)، يمكننا القول بأنه بعد ترتيب درجات الأفراد قد تم استخدام اختبار كروسكال - وليز لتقييم الفروق بين المجموعات الثلاثة. وقد جاءت النتيجة لتشير إلى أنه لا توجد فروق دالة بين هذه المجموعات.

$$^{(12)}\ H: 3.785(df2, N=15)P > .05$$

رابعاً: اختبار فريدمان: كبديل لمقاييس أنوفا المتكررة The Friedman Test:

يستخدم اختبار فريدمان لتقييم الفروق بين ثلاث أو أربع مجموعات مستخدماً بيانات لتصميم مقاييس متكررة.

ويعتبر اختبار فريدمان بديلاً لمقاييس انوفا المتكررة التي تحدثنا عنها في الفصول السابقة. ومع أن تحليل أنوفا يتطلب درجات عددية Numerical Scores التي يمكن استخدامها في حساب المتوسطات والتباين، إلا أن اختبار فريدمان يتطلب ببساطة أن يكون الباحث قادراً على ترتيب المشاركين على المتغير الذي يتم قياسه. كذلك نجد أن مقياس فريدمان مشابه لاختبار ولكوكسن. حيث إن الأخير - كما أشرنا سابقاً - يتعلق بمقارنة حالتين فقط. في حين أن اختبار فريدمان يستخدم للمقارنة بين ثلاث أو أكثر من الحالات.

مستوى البيانات المطلوبة لاختبار فريدمان:

لإجراء اختبار فريدمان، يتطلب الأمر وجود عينة واحدة One Sample لكل فرد مشارك في كل حالات المعالجة المختلفة. ويتطلب هذا الاختبار ترتيب كل حالات المعالجة لكل فرد مشارك في الدراسة. كما يمكن استخدام اختبار فريدمان إذا كانت البيانات الأصلية تحتوي على درجات عددية، ومع ذلك ينبغي تحويل هذه الدرجات العددية إلى رتب قبل إجراء هذا الاختبار.

والمثال التالي يبين طريقة استخدام هذا الاختبار.

مثال: عينة تتكون من خمسة مشاركين N = 5، تم اختبارهم تحت أربعة أنواع من

الدواء. وقد كانت الدرجات الأصلية تحتوي على درجات عددية (انظر الجدول "أ"). ولكي يطبق اختبار فريدمان فقد حُولت هذه الدرجات العددية إلى رتب 1، 2، 3 و 4 لتطابق حجم الدرجات الأصلية. فعلى سبيل المثال، أن المشارك رقم (1) لديه درجات 3، 4، 6، 7. ولهذا المشارك X = 3 وهي أصغر درجة. وبالتالي كان ترتيبه (1)؛ و X = 4 هي الدرجة الثانية من حيث الصغر؛ وبالتالي كان ترتيبه (2)، أما الدرجتان الأخريان فقط تحصلتا على ترتيب (3) و (4). في حين أن المشارك رقم (2) لديه درجات متعادلة لِـ: X = 3 للدواء (أ) والدواء (ب)؛ وبالتالي قد تحصلا على ترتيب (2.5). (2+3 = 2/5 تساوي 2.5). إن مجموعة الرتب كاملة يبينها الجزء من الجدول (20 – ب).

(أ): جدول (20 ـ أ) الدرجات الأصلية للدواء

المشارك	دواء بدون مفعول (مهدئ للمريض)	عقار (أ)	عقار (ب)	عقار (ج)
1	3	4	6	7
2	0	3	3	6
3	2	1	4	5
4	0	1	3	4
5	0	1	4	3

(ب): جدول (20 ـ ب) ترتيب حالات المعالجة بشكل منظم لكل مشارك

المشارك	دواء بدون مفعول (مهدئ للمريض)	عقار له مفعول (أ)	عقار له مفعول (ب)	عقار له مفعول (ج)
1	1	2	3	4
2	1	2.5	2.5	4
3	2	1	3	4
4	1	2	3	4
5	1	2	4	3
	$R_1 = 6$	$R_2 = 9.5$	$R_3 = 15.5$	$R_4 = 19$

المصدر: .Frederick J. GRavetter and Larry B. wallnau , op.cit , p.685

الفرض الصفري لاختبار فريدمان:[13]

بشكل عام، فإن الفرض الصفري لاختبار فريدمان يصاغ بالشكل التالي:

لا توجد فروق بين حالات المعالجة الخاضعة للمقارنة. فإذا كان الفرض الصفري فرضاً صحيحاً، فإن الرتب يجب أن تتوزع بشكل عشوائي؛ وبالتالي لا توجد أية نزعة بأن هذه الرتب لإحدى المعالجات تكون بشكل نظامي أعلى (أو أقل) من الرتب في المعالجات الأخرى. وعليه، فإن الفروض المتعلقة باختبار فريدمان يمكن أن تصاغ بالشكل التالي:

Ho: لا يوجد فرق بين المعالجات. وعليه فإن الرتب في إحدى حالات المعالجة لا يجب أن تكون أعلى أو أقل من الرتب في أية معالجة أخرى.

Hi: يوجد فروق بين المعالجات، وعليه، فإن الرتب على الأقل في إحدى حالات المعالجة يجب أن تكون بشكل منظم أعلى أو أقل من الرتب في معالجة أخرى.

معادلة وحساب اختبار فريدمان:

لحساب اختبار فريدمان ينبغي للباحث اتباع الخطوات التالية:

1- ترتيب درجة كل مشارك من المشاركين الخمسة.

2- تجمع الرتب لكل فرد مشارك كما هو مبين في السطر الأخير من الجدول. فإذا كان الفرض الصفري فرضاً صحيحاً، فإن هذه المجاميع يجب أن تكون تقريباً بنفس الحجم. لأنه لا أحد من هؤلاء المشاركين لديه ترتيب أعلى أو أدنى من المشارك الآخر. وعلى الجانب الآخر، إذا ما وجد فرق ثابت بين هؤلاء المشاركين، إذاً على الأقل واحد من هذه المجاميع يجب أن تكون لافتة للنظر بأنها أكبر (أو أصغر) من مجموع مشارك أخرى.

ويشار إلى مجموع الرتب لكل مشارك بالرمز R_1، R_2 وهكذا... ففي الجدول (ب) نجد أن المعالجة الأولى (الدواء المهدئ Placebo) يصل مجموعها إلى $R_1 = 6$ في حين يصل مجموع المعالجة الثانية $R_2 = 9.5$، والمعالجة الثالثة $R_3 = 15.5$، في حين يصل مجموع المعالجة الأخيرة إلى: $R_4 = 19$.

أما باقي الرموز المتعلقة باختبار فريدمان تحتوي على n و K والتي تقابل عدد الأفراد في العينة وعدد حالات المعالجة على التوالي. ففي البيانات الواردة في جدول (ب) فإن: N = 5 و K = 4.

ولما كان اختبار فريدمان اختباراً لتقييم الفروق بين المعالجات من خلال حساب إحصائي الاختبار التالي:

$$X_r^2 = \frac{12}{nk(k+1)}\sum R^2 - 3n(k+1)$$

لا حظ أن الإحصاء تم تعيّنه كمربع كاي (X^2) مقروناً برمز سفلي (r) ويقابل إحصاء مربع كاي (X^2) للرتب.

إن إحصاء مربع كاي لديه درجات حرية تحدد من خلال: df=k-1 ويقيِّم من خلال القيم الحرجة في توزيعات مربع كاي. وباستخدامنا للبيانات الواردة في الجدول رقم (ب)، فإننا سنتحصل على الإحصاء التالي:

$$X_r^2 = \frac{12}{nk(k+1)}\sum R^2 - 3n(k+1)$$

$$X_r^2 = \frac{12}{5(4)(5)}\left((6)^2 + (9.5)^2 + (15.5)^2 + (19)^2\right) - 3(5)(5)$$

$$= \frac{12}{100}(36 + 90.25 + 240.25 + 361) - 75$$

$$= 0.12\ (727.5) - 75$$

$$= 12.3$$

وبدرجة حرية df=k-1=3، وقيمة حرجة لمربع كاي تساوي 9.35. عليه فإن القرار الذي نصل إليه هو رفض الفرض الصفري، وبالتالي نصل إلى نتيجة مفادها أنه توجد فروق ذات دلالة إحصائية بين أحوال المعالجات الأربع.

لاحظ أنه في اختبار فريدمان ببساطة يمكننا القول بأنه توجد فروق دالة، في حين يمكننا القول عند التعامل مع اختبار أنوفا أنه توجد دلالة في فروق المتوسط. ولتفسير

نتائج اختبار فريدمان - مثله مثل كل الاختبارات ذات البيانات الترتيبية - لا يوجد شكل عام لإعداد تقرير حول النتائج المتحصل عليها من خلال هذا الاختبار؛ ومع ذلك، فإنه في إعداد التقرير المتعلق به ينبغي أن يحتوي على: القيمة المتحصل عليها بالنسبة لإحصاء مربع كاي، إضافة إلى القيم بالنسبة لدرجة الحرية df و N و P. بالنسبة للبيانات التي بين أيدينا التي تم البرهنة بها على هذا الاختبار، فالتقرير المعد ينبغي أن يدور حول النتائج التالية:

- بعد ترتيب البيانات الأصلية، فقد تم استخدام اختبار فريدمان لتقييم الفروق بين أربع حالات معالجة. وقد أشارت النتيجة أن هناك فروقاً دالة:

$$X_r^2 = (12.3(df = 3, n = 5), P < .05)$$

جدول (20 ـ 3) ملخص: مقارنة بين هذه المقاييس الأربعة

الاختبار	نمط العلاقات المختبرة	مستوى البيانات المطلوبة	تصميم الدراسة
مان-وتني U	تقييم الفرق بين حالتين	بيانات ترتيبية	بين الموضوعات: • عينات مستقلة • موضوع فردي
ولكوكسن (T)	تقييم الفرق بين حالتين	بيانات ترتيبية	داخل الموضوعات: • المقاييس المتكررة • الأزواج المتماثلة
كروسكال-وليز (H)	تقييم الفروق بين ثلاث حالات أو أكثر	بيانات ترتيبية	بين الموضوعات: • تصميمات غير مترابطة • (موضوعات مستقلة)
فريدمان	تقييم الفروق بين مجموعتين أو أكثر	بيانات ترتيبية	• مقاييس متكررة / أزواج متماثلة

الإرشادات العامة لاستخدام SPSS لحساب اختبار: مان وتني U، ولكوكسن T، كروسكال - وليز وفريدمان.

أولا: اختبار مان وتني U:

الإجراء المتبع:

بعد إدخال البيانات Data Entry يقوم الباحث بتحليلها وفقاً للإجراء التالي:

1- في القائمة المعروضة في الجزء العلوي من الشاشة:انقر فوق Analyze ⟸ Non-Parametric tests ⟸ و انقر على Independent 2 Samples.

2- التأكد من أن صندوق Mann-Whitney قد تم تحديده.

3- انقر فوق المتغير التابع، وانقله إلى مربع Test Variable List.

4- انقر فوق المتغير المستقل، وانقله إلى مربع Grouping Variable.

5- انقر فوق زر Define Groups واكتب قيمة Group 1، وقيمة Group 2، وأدخل قيمة 1 و 2 إلى مجموعة الصناديق الملائمة.

6- انقر فوق Contiue.

7- انقر فوق OK.

المخرجات Spss output:

يُنتج هذا الإجراء جدولين يعرض الأول عدد الدرجات، متوسط الرتب ومجموع الرتب لكل مجموعة. أما الجدول الثاني فيقوم بعرض نتائج الاختبار مشتملاً على قيمة U، ودرجة Z التقريبية، وقيمة مستوى الدلالة لكل من مان وتني، و Z (Asymp. sig 2-tailed). لاحظ أنه إذا كان حجم العينة أكبر من 30، فإن برنامج SPSS سيولد قيمة اختبار Z-approximation. فإذا كانت القيمة المحسوبة لـِ U من بيانات العينة أصغر أو مساوية

للقيمة الجدولية، فالباحث إذاً يمكنه الوصول إلى نتيجة مفادها أن الفرق بين المجموعتين ذو دلالة إحصائية.

ثانيا: اختبار ولكوكسن Wilcoxon test:

الإجراء المتبع:

بعد إدخال البيانات Data Entry يقوم الباحث بتحليلها وفقاً للإجراء التالي:

1- في القائمة الموجودة في الجزء العلوي من الشاشة انقر على:
Analyze ⟸ Non-Parametric tests ⟸ وانقر على Related Samples2

2- التأكد من أن صندوق Wilcoxon test قد تم تحديده Test type.

3- انقر فوق المتغيرات التي تمثل الدرجات في كل من 1، 2، وانقلها إلى مربع Test Paired List.

4- انقر فوق OK.

المخرجات Spss output:

إن الأمرين الأساسيين اللَّذَيْن ينبغي على الباحث التركيز عليهما من خلال الجدولين (جدول يعرض العدد "N"، ومتوسط الرتب، ومجموع الرتب - السالبة والموجبة -، والجدول الثاني يعرض نتائج الاختبار مشتملاً على القيمة التقريبية لدرجة Z، ومستوى الدلالة) هما قيمة Z، ومستويات الدلالة المرتبطة بها والتي يعبر عنها بِـ: Asymp. Sig. (2-tailed). فإذا كان مستوى الدلالة أقل أو مساوياً لِـ 05. (05. ، 01. ، 0.001)، فإنه بالتالي يصل إلى قرار مفاده بأن هناك فرقاً ذا دلالة إحصائية.

ثالثا: اختبار كروسكال ـ وليز Kruskal - Wallis:

الإجراء المتبع:

بعد إدخال البيانات Data Entry يقوم الباحث بتحليلها وفقاً للإجراء التالي:

1- في القائمة الموجودة في الجزء العلوي من الشاشة: انقر فوق:

Analyze ⟸ Non-Parametric tests وانقر K Independent Samples .

2- التأكد من أن مربع Kruskal - Wallis قد تم تحديده في جزء Test type.

3- انقر فوق المتغير التابع وانقله إلى مربع Test Variable List.

4- انقر فوق المتغير المستقل، وانقله إلى مربع Grouping Variable.

5- انقر فوق زر Define Range وأدخل القيم Minimum و Maximum في مربع Grouping Variable (على سبيل المثال) إذا استخدمت 1، 2 و 3 لبيان المعالجات الثلاث، أدخل Minimum of 1 و Maximum of 3.

6- انقر فوق Contiue.

7- انقر فوق OK.

المخرجات Spss output:

في هذا الاختبار ينبغي على الباحث التركيز على الجدول الثاني من المخرجات والذي يحتوي على قيمة X^2، ودرجة الحرية، ومستوى الدلالة. Asymp.sig. ومن هنا يتوجب على الباحث أن يكون قراره كالتالي:

إذا كان مستوى الدلالة أقل من 0.05، فإن ذلك يعني أن هناك فرقاً دالاً إحصائياً.

رابعا: اختبار Friedman:

الإجراء المتبع:

بعد إدخال البيانات Data Entry يقوم الباحث بتحليلها وفقاً للإجراء التالي:

1- في القائمة الموجودة في الجزء العلوي من الشاشة: انقر فوق

Analyze ⟸ Non-Parametric tests انقر على K Related Samples

2- تأكد من اختبار Friedman test قد تم تحديده في جزء Test type.

3- انقر فوق المتغيرات التي تمثل موضوعات القياس 1، 2، 3.

4- انقر فوق OK.

المخرجات Spss output:

تحتوي المخرجات لهذا الاختبار على جدولين. الجدول الأول يبين متوسط الرتب لكل حالة معالجة. أما الجدول الثاني فيشتمل على نتائج الاختبار: قيمة مربع كاي (X^2)، درجة الحرية (df) ومستوى الدلالة (α) قيمة الفا للاختبار.

أسئلة للمراجعة:

1- رتب الدرجات التالية متى كان ذلك ضرورياً:

14، 3، 4، 0، 3، 5، 14، 3.

2- تجربة استخدمت عينة مؤلفة من N = 25 في العينة الأولى و10 = N في العينة الثانية. ولكي نجري اختبار مان – وتني U لـ: 50 = U مفترضاً أن هذا العدد هو أصغر قيمة U. ما هي قيمة U العينة الأخرى.

3- في دراسة أوضحت أن درجات عينة الأولاد (4) ودرجات عينة البنات (9) هي كالتالي:

درجة الأولاد: 21، 14، 17، 8

درجة البنات: 18، 25، 23، 21، 34، 28، 32، 30، 13

المطلوب: إجراء اختبار مان – وتني.

4- طبقاً لجدول مان – وتني، فإن قيمة 30 = U ، وهي دالة على مستوى 05. لكلتا العينتين 11 = N. وإذا ما تم استخدام هذه القيمة في التوزيع القريب من الطبيعي Normal approximation، هل تولد لنا درجة Z في المنطقة الحرجة؟

5- أراد طبيب أن يختبر فعالية دواء جديد لالتهاب المفاصل وذلك لقياس قوة قبض اليد لدى المرضى قبل وبعد استخدام هذا الدواء. وقد كانت فروق الدرجات لعشرين من المرضى كالتالي:

+ 41 ،−2، + 25 ، −8 (لا تغير) 0 ، + 14 + 34 ، + 16 + 46 ، + 3

وتمثل كل درجة الفرق في قوة قبض اليد. وتشير (+) إلى قوة قبض اليد بعد تناول الدواء.

المطلوب: استخدام اختبار T لتحديد ما إذا كانت هذه البيانات تقدم دليلاً كافياً مفاده أن الدواء الجديد لديه تأثير دال.

اختبر هذه البيانات على مستوى $\alpha = 0.05$.

6- البيانات التالية تحتوي على عينة مؤلفة من 7 أشخاص N = 7 وتم اختبار كل فرد من هؤلاء في وضعين مختلفين من المعالجة. والبيانات لهذه العينة يظهرها الجدول التالي:

المشاركون	المعالجة الأولى (1)	المعالجة الثانية (2)	الفرق
1	8	24	+ 16
2	12	10	− 2
3	15	19	+ 4
4	31	52	+ 21
5	26	20	− 6
6	32	40	+ 8
7	19	29	+ 10

المطلوب: إجراء اختبار ولكوكسن من خلال الخطوات التي تعلمتها من مطالعتك لهذا الفصل.

7- البيانات التالية تظهر نتائج جاءت من دراسة استخدمت عينات مستقلة لغرض المقارنة في أوضاع من المعالجات. بعد المعالجة فقد تم ترتيب كل الأفراد في العينات الثلاث. وأن حجم الرتب لـِ: T تم حسابه لكل معالجة.

1	2	3
8	1	11
2	7	5
4	6	9
10	3	12
14	13	15
$T_1 = 38$	$T_2 = 30$	$T_3 = 52$
$n_1 = 5$	$n_2 = 5$	$n_3 = 5$

المطلوب: إجراء اختبار كروسكال – وليز من خلال الخطوات التي تعلمتها من خلال هذا الاختبار؟

8- في دراسة أوضحت نتائجها أن طلاب المدارس الثانوية الذين هم بشكل منتظم يشاهدون بعض الرسوم المتحركة يسجلون درجات عالية مقارنة بنظرائهم الذين لم يشاهدوا هذه الرسوم. وإن نتيجة اختبار T للمقاييس المستقلة، تشير إلى فرق دال بين المجموعتين. ولإعادة البيانات هنا:

أ- هل اختبار مان - وتني، يشير أيضاً إلى فرق دال؟ (لاستخدام مستوى دلالة $\alpha = 0.05$).

ب- عادة ما يجرى اختبار كروسكال - وليز لتقييم الفرق بين ثلاث مجموعات من العينات أو أكثر. ومع هذا فإنه بالإمكان استخدام هذا الاختبار للمقارنة بين مجموعتين من العينات. احسب اختبار كروسكال - وليز مستخدماً درجة دلالة $\alpha = 0.05$ لتقييم الفرق بين هاتين المجموعتين من الطلاب. هل النتيجة التي ستحصل عليها متسقة مع النتيجة المتحصل عليها من خلال اختبار مان - وتني.

9- مسح اجتماعي لعينة مؤلفة من 7 من الطلاب الجامعيين يطلب منهم أن يرتبوا بشكل منظم الاحتمالات الثلاثة التالية فيما يتعلق بأقرب الأشخاص في علاقاتهم الاجتماعية:

أ- صديق من نفس الجنس.

ب- صديق من الجنس المقابل.

ج- عضو من العائلة من أي من الجنسين.

والبيانات التالية تظهر في الجدول التالي:

الطالب	نفس الجنس	الجنس المقابل	أحد أعضاء العائلة
1	1	3	2
2	1	2	3
3	2	1	3
4	2	3	1
5	2	1	3
6	3	2	1
7	1	2	3
المجموع	12	14	16

المطلوب: إجراء اختبار فريد مان مستخدماً الإجراءات التي تعلمتها من خلال مطالعتك لهذا الفصل (مستوى الدلالة $\alpha = .05$).

10- نفترض أننا نرغب في دراسة المقاييس المتكررة لمقارنة ثلاث حالات من المعالجة مستخدماً نفس العينة المؤلفة من $N = 6$ مشاركين في كل معالجة. يطلب منك التفكير في احتمالين من النتائج المتطرفة لهذه الدراسة.

أ- تخيل أنه لا يوجد تداخل Overlap بين المجموعات الثلاث، وتحديداً. افترض أن كل المشاركين الستة تم ترتيبهم الأول في المعالجة الأولى، والثاني في المعالجة الثانية، والثالث في المعالجة الثالثة، في هذه الحالة، يكون حجم الترتيب للمعالجات الثلاثة هو $R_1=6$, $R_2=12$, $R_3=18$. السؤال المطروح للإجابة عليه هو: ما هي النتيجة المتوقعة لهذه البيانات من خلال اختبار فريد مان. احسب قيمة X^2.

ب- تخيل الآن أنه لا توجد فروق بين المجموعات الثلاثة من المعالجات. على سبيل المثال، فأن المشاركين الثلاث الأول لديهم ترتيب 1، 2 و 3. عبر المعالجات الثلاثة. وأن آخر المشاركين الثلاثة لديهم رتب 3، 2 و 1. في هذه الحالة فإن كل معالجة لديها نفس القيمة $R = 12$ للمجموع الكلي للرتب.

مرة ثانية: ما هي النتيجة المتوقعة لاختبار فريدمان، ثم أحسب قيمة X^2.

الهوامش والمصادر:

أولا: الهوامش:

1- Frederick J. Gravetter and Larry B. Wallnau , Statistics for The Behavioral Sciencies , 8th ed. , Wadsworth Cengage Learning , 2010 , P. 666.

2- Ibid , PP. 667 - 668.

3- Ibid , PP. 669 - 673.

4- Ibid , P. 672.

5- Ibid , P. 672.

6- Ibid , PP. 675 - 676.

7- Ibid , P. 676.

8- Ibid , P. 676.

9- Ibid , PP. 676 - 677.

10- Ibid , PP. 677 - 678.

11- Ibid , PP. 679 - 680.

12- Ibid , P. 683.

13- Ibid , PP. 684 - 687.

ثانيا: المصادر:

1- Frederick J. Gravetter and Larry B. Wallnau , Statistics for The Behavioral Sciencies , 8th ed. , Wadsworth Cengage Learning , 2010.

2- George Argyrous , Statistics for Social and Health Research With a Guide to SPSS , Sage Publications , London , 2001.

3- George DiEkhoff, Statistics for the social and Behavioral Scienaces: Univariate, Bivariate, Multivariate , MCGRAW-Hill Companies INC. USA , 1992.

4- جولي بالانت (ت) خالد العامري، التحليل الإحصائي باستخدام SPSS، دار الفاروق للنشر والتوزيع، مصر، 2009 م.

الملاحـق

- ملحق (1) الجداول الإحصائية
- ملحق (2) مسرد لبعض المفاهيم الإحصائية الواردة في متن المصنف
- ملحق (3) منظم إحصائي
- ملحق (4) بعض المعادلات الأساسية الواردة في الكتاب

المصادر:

- أولاً: باللغة العربية
- ثانيا: باللغة الإنجليزية

ملحق (1): الجداول الإحصائية

- جدول 1/1 : المساحة تحت المنحنى الطبيعي المعياري .
- جدول 2/1 : توزيعات t .
- جدول 3/1 : توزيعات F .
- جدول 4/1 : القيم الحرجة لارتباط بيرسون (r) .
- جدول 5/1 : القيم الحرجة لارتباط سبيرمان (rs) .
- جدول 6/1 : توزيعات مربع كاي (X2) .
- جدول 7/1 : القيم الحرجة لاختبار مان-وتني U .
- جدول 8/1 : القيم الحرجة لاختبار ولكوكسن T .

ملحق (2): مسرد المفاهيم الإحصائية الواردة في متن المصنف

ملحق (3) منظم إحصائي .

ملحق (4) بعض المعادلات الأساسية الواردة في الكتاب .

المصادر

- أولا : باللغة العربية .
- ثانيا : باللغة الإنجليزية .

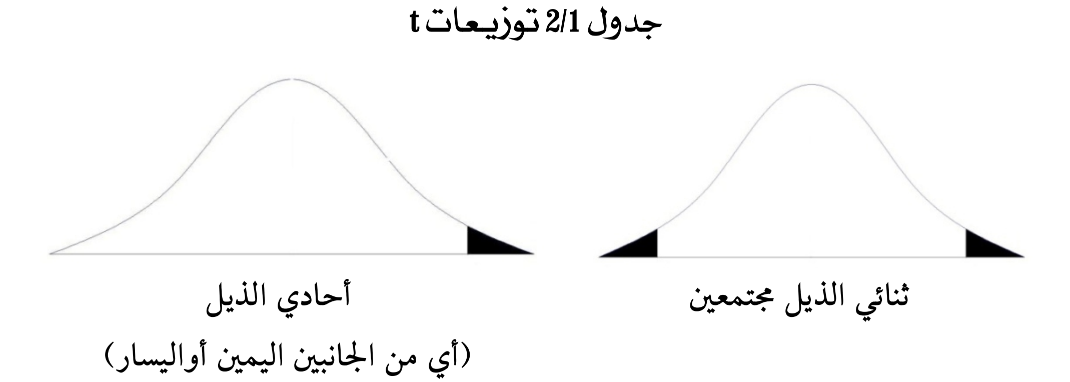
ملحق (1): الجداول الإحصائية

جدول 1/1 المساحة تحت المنحنــى الطبيـعي المعياري

Z	المساحة تحت المنحنى بين كلتا النقطتين	المساحة تحت المنحنى أبعد من النقطتين	المساحة تحت المنحنى أبعد من نقطة واحدة
±0.1	0.080	0.920	0.4600
±0.2	0.159	0.841	0.4205
±0.3	0.236	0.764	0.3820
±0.4	0.311	0.689	0.3445
±0.5	0.383	0.617	0.3085
±0.6	0.451	0.549	0.2745
±0.7	0.516	0.484	0.2420
±0.8	0.576	0.424	0.2120
±0.9	0.632	0.368	0.1840
±1	0.683	0.317	0.1585
±1.1	0.729	0.271	0.1355
±1.2	0.770	0.230	0.1150
±1.3	0.806	0.194	0.0970
±1.4	0.838	0.162	0.0810
±1.5	0.866	0.134	0.0670
±1.6	0.890	0.110	0.0550
±1.645	0.900	0.100	0.0500
±1.7	0.911	0.089	0.0445
±1.8	0.928	0.072	0.0360
±1.9	0.943	0.057	0.0290
±1.96	0.950	0.050	0.0250
±2	0.954	0.046	0.0230
±2.1	0.964	0.036	0.0180

Z	المساحة تحت المنحنى بين كلتا النقطتين	المساحة تحت المنحنى أبعد من النقطتين	المساحة تحت المنحنى أبعد من نقطة واحدة
±2.2	0.972	0.028	0.0140
±2.3	0.979	0.021	0.0105
±2.33	0.980	0.020	0.0100
±2.4	0.984	0.016	0.0080
±2.5	0.988	0.012	0.0060
±2.58	0.990	0.010	0.0050
±2.6	0.991	0.009	0.0045
±2.7	0.993	0.007	0.0035
±2.8	0.995	0.005	0.0025
±2.9	0.996	0.004	0.0020
±3	0.997	0.003	0.0015
±3.1	0.998	0.002	0.0001
±3.2	0.9986	0.0014	0.0007
±3.3	0.9990	0.0010	0.0005
±3.4	0.9993	0.0007	0.0003
±3.5	0.9995	0.0005	0.00025
±3.6	0.9997	0.0003	0.00015
±3.7	0.9998	0.0002	0.00010
±3.8	0.9986	0.00014	0.00007
±3.9	0.99990	0.00010	0.00005
±4	>0.99990	<0.00010	<0.00005

جدول 2/1 توزيـعات t



	أحادي الجانب					
	0.25	0.10	0.05	0.25	0.01	0.005
	ثنائي الجانب					
Df	0.50	0.20	0.10	0.05	0.02	0.01
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707
7	0.711	1.415	1.895	2.365	2.998	3.499
8	0.706	1.397	1.860	2.306	2.896	3.355
9	0.703	1.383	1.833	2.262	2.821	3.250
10	0.700	1.372	1.812	2.228	2.764	3.169
11	0.697	1.363	1.796	2.201	2.718	3.106
12	0.695	1.356	1.782	2.179	2.681	3.055
13	0.694	1.350	1.771	2.160	2.650	3.012
14	0.692	1.345	1.761	2.145	2.624	2.977
15	0.691	1.341	1.753	2.131	2.602	2.947
16	0.690	1.337	1.746	2.120	2.583	2.921
17	0.689	1.333	1.740	2.110	2.567	2.898
18	0.688	1.330	1.734	2.101	2.552	2.878
19	0.688	1.328	1.729	2.093	2.539	2.861
20	0.687	1.325	1.725	2.086	2.528	2.845
21	0.686	1.323	1.721	2.080	2.518	2.831

	أحادي الجانب					
	0.25	0.10	0.05	0.25	0.01	0.005
	ثنائي الجانب					
Df	0.50	0.20	0.10	0.05	0.02	0.01
22	0.686	1.321	1.717	2.074	2.508	2.819
23	0.685	1.319	1.714	2.069	2.500	2.807
24	0.685	1.318	1.711	2.064	2.492	2.797
25	0.684	1.316	1.708	2.060	2.485	2.787
26	0.684	1.315	1.706	2.56	2.479	2.779
27	0.684	1.314	1.703	2.052	2.473	2.771
28	0.683	1.313	1.701	2.048	2.469	2.763
29	0.683	1.311	1.699	0.045	2.462	2.756
30	0.683	1.310	1.697	2.042	2.457	2.750
40	0.681	1.303	1.684	2.021	2.423	2.704
60	0.679	1.296	1.671	2.000	2.390	2.660
120	0.677	1.289	1.658	1.980	2.358	2.617
∞	0.674	1.282	1.645	1.960	2.326	2.576

جدول 3/1 توزيـعات F

مستوى الدلالة 0.05 = ∝

n_1 / n_2	1	2	3	4	5	6	8	12	24	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	238.9	243.9	249.0	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.41	19.45	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.84	8.74	8.64	8.53
4	7.71	6.94	6.59	6.59	6.39	6.26	6.16	6.04	5.91	5.77
5	6.16	5.79	5.41	5.19	5.05	4.95	4.82	4.68	4.53	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.00	3.84	3.67
7	5.99	4.74	4.35	4.12	3.97	3.87	3.73	3.57	3.41	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.28	3.12	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.07	2.90	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.91	2.74	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	2.95	2.79	2.61	2.40
12	4.75	3.88	3.49	3.26	3.11	3.00	2.85	2.69	2.50	2.30
13	4.67	3.80	3.41	3.18	3.02	2.92	2.77	2.60	2.42	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.53	2.35	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.48	2.29	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.42	2.24	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.55	2.38	2.19	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.34	2.15	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.31	2.11	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.28	2.08	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.42	2.25	2.05	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.40	2.23	2.03	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.38	2.20	2.00	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.36	2.18	1.98	1.73
25	4.24	3.38	2.99	2.76	2.60	2.49	2.34	2.16	1.96	1.71
26	4.22	3.37	2.98	2.74	2.59	2.47	2.32	2.15	1.95	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.30	2.13	1.93	1.67
28	4.20	3.34	2.95	2.71	2.56	2.44	2.29	2.12	1.91	1.65
29	4.18	3.33	2.93	2.70	2.54	2.43	2.28	2.10	1.90	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.09	1.89	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.00	1.79	1.51
60	4.00	3.15	2.76	2.52	2.37	2.25	2.10	1.92	1.70	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.02	1.83	1.61	1.25
∞	3.84	2.99	2.60	2.37	2.21	2.09	1.94	1.75	1.52	1.00

جدول 3/1 توزيعات F
مستوى الدلالة 0.01 = α

n_1 / n_2	1	2	3	4	5	6	8	12	24	∞
1	4052	4999	5403	5625	5764	5854	5981	6106	6234	6366
2	98.49	99.01	99.17	99.25	99.30	99.33	99.36	99.42	99.46	99.50
3	34.12	30.81	29.46	28.71	28.24	27.91	27.49	27.05	26.60	26.12
4	21.20	18.00	16.69	15.98	15.52	15.21	14.80	14.37	13.93	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.27	9.89	9.47	9.02
6	13.74	10.92	9.78	9.15	8.75	8.47	8.10	7.72	7.31	6.88
7	12.25	9.55	8.45	7.85	7.46	7.15	6.84	6.47	6.07	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.03	5.67	5.28	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.47	5.11	4.73	4.31
10	10.04	7.56	6.55	5.99	5.54	5.39	5.06	4.71	4.33	3.91
11	9.65	7.20	6.22	5.67	5.32	5.07	4.74	4.40	4.02	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.50	4.16	3.78	3.36
13	9.07	6.70	5.74	5.20	4.86	4.62	4.30	3.96	3.59	3.16
14	8.86	6.51	5.56	5.03	4.69	4.46	4.14	3.80	3.43	3.00
15	9.68	6.36	5.42	4.89	4.56	4.32	4.00	3.67	3.29	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	3.89	3.55	3.18	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.79	3.45	3.08	2.65
18	8.28	6.01	5.09	4.58	4.25	4.01	3.71	3.37	3.00	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.63	3.30	2.92	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.56	3.23	2.86	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.51	3.17	2.80	2.36
22	7.94	5.72	4.82	4.31	3.99	3.76	3.45	3.12	2.75	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.41	3.07	2.70	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.36	3.03	2.66	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.32	2.99	2.62	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.29	2.96	2.58	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.26	2.93	2.55	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.23	2.90	2.52	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.20	2.87	2.49	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.17	2.84	2.47	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	2.99	2.66	2.29	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.82	2.50	2.12	1.60
120	6.85	4.75	3.95	3.48	3.17	2.96	2.66	2.34	1.95	1.38
∞	6.64	4.60	3.78	3.32	3.02	2.80	2.51	2.18	1.79	1.00

جدول 4/1 القيم الحرجة لارتباط بيرسون (r)*

* لكي يكون ارتباط بيرسون(r) ارتباطاً دالاً ، لابد أن تكون قيمة (r) أكبر أو مساوية للقيمة الحرجة في الجدول .

Df=N-2	مستوى الدلالة لاختبار أحادي الجانب			
	.05	.025	.01	.005
	مستوى الدلالة لاختبار ثنائي الجانب			
	.10	.05	.02	.01
1	.988	.997	.9995	.9999
2	.900	.950	.980	.990
3	.805	.878	.934	.999
4	.729	.811	.882	.917
5	.665	.754	.833	.874
6	.622	.707	.789	.834
7	.582	.666	.750	.798
8	.549	.632	.716	.765
9	.521	.602	.685	.735
10	.497	.576	.658	.708
11	.476	.553	.634	.684
12	.458	.532	.612	.661
13	.441	.514	.592	.641
14	.426	.497	.574	.623
15	.412	.482	.558	.606
16	.400	.468	.542	.590
17	.389	.456	.528	.575
18	.378	.444	.516	.561
19	.369	.433	.503	.549
20	.360	.423	.492	.537
21	.352	.413	.482	.526
22	.344	.404	.472	.515
23	.337	.396	.462	.505
24	.330	.388	.453	.496
25	.323	.381	.445	.487
26	.317	.374	.437	.479
27	.311	.367	.430	.471
28	.306	.361	.423	.463
29	.301	.355	.416	.456
30	.296	.349	.409	.449
35	.275	.325	.381	.418
40	.257	.304	.358	.393
45	.243	.288	.338	.372
50	.231	.273	.322	.354
60	.211	.250	.295	.325
70	.195	.232	.274	.302
80	.183	.217	.256	.283
90	.173	.205	.242	.267
100	.164	.195	.230	254

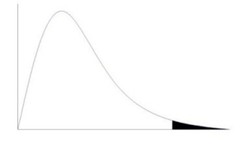
جدول 5/1 القيم الحرجة لارتباط سبيرمان *

*** لكي يكون ارتباط سبيرمان ارتباطاً دالاً ، لابد أن تكون قيمة r_s أكبر أو مساوية للقيمة الحرجة في الجدول .**

	مستوى الدلالة لاختبار أحادي الجانب			
	.05	.025	.01	.005
N	مستوى الدلالة لاختبار ثنائي الجانب			
	.10	.05	.02	.01
4	1.000			
5	0.900	1.000	1.000	
6	0.829	0.886	0.943	1.000
7	0.714	0.786	0.893	0.929
8	0.643	0.738	0.833	0.881
9	0.600	0.700	0.783	0.833
10	0.564	0.648	0.754	0.794
11	0.536	0.618	0.709	0.755
12	0.503	0.587	0.671	0.727
13	0.484	0.560	0.648	0.703
14	0.464	0.538	0.622	0.675
15	0.443	0.521	0.604	0.654
16	0.429	0.503	0.582	0.635
17	0.414	0.485	0.566	0.615
18	0.401	0.472	0.550	0.600
19	0.391	0.460	0.535	0.584
20	0.380	0.447	0.520	0.570
21	0.370	0.435	0.508	0.556
22	0.361	0.425	0.496	0.544
23	0.353	0.415	0.486	0.532
24	0.344	0.406	0.476	0.521
25	0.337	0.398	0.466	0.511
26	0.331	0.390	0.457	0.501
27	0.324	0.382	0.448	0.491
28	0.317	0.375	0.440	0.483
29	0.312	0.368	0.433	0.475
30	0.306	0.362	0.425	0.467
35	0.283	0.335	0.394	0.433
40	0.264	0.313	0.368	0.405
45	0.248	0.294	0.347	0.382
50	0.235	0.279	0.329	0.363
60	0.214	0.255	0.300	0.331
70	0.190	0.235	0.278	0.307
80	0.185	0.220	0.260	0.287
90	0.174	0.207	0.245	0.271
100	0.165	0.197	0.233	0.257

جدول 6/1 توزيـعات مربـع كـاي X^2 *

* مدخلات الجدول تشير إلى القيم الحرجة لمربع كاي X^2 .



X^2 الحرجة

Df	النسبة في المنطقة الحرجة *Proportion in Critical Region*				
	0.10	0.05	0.025	0.01	0.005
1	2.71	3.84	5.02	6.63	7.88
2	4.61	5.99	7.38	0.21	10.60
3	6.25	7.81	9.35	11.34	12.84
4	7.78	9.49	11.14	13.28	14.86
5	9.24	11.07	12.83	15.09	16.75
6	10.64	12.59	14.45	16.81	18.55
7	12.02	14.07	16.01	18.48	20.28
8	13.36	15.51	17.53	20.09	21.96
9	14.68	16.92	19.02	21.67	23.59
10	15.99	18.31	20.48	23.21	25.19
11	17.28	19.68	21.92	24.72	26.76
12	18.55	21.03	23.34	26.22	28.30
13	19.81	22.36	24.74	27.69	29.82
14	21.06	23.68	26.12	29.14	31.32
15	22.31	25.00	27.49	30.58	32.80
16	23.54	26.30	28.85	32.00	34.27
17	24.77	27.59	30.19	33.41	35.72
18	25.99	28.87	31.53	34.81	37.16
19	27.20	30.14	32.85	36.19	38.58
20	28.41	31.41	34.17	37.57	40.00
21	29.62	32.67	35.48	38.93	41.40
22	30.81	33.92	36.78	40.29	42.80
23	32.01	35.17	38.08	41.64	44.18
24	33.20	36.42	39.36	42.98	45.56
25	34.38	37.65	40.65	44.31	46.93
26	35.56	38.89	41.92	45.64	48.29
27	36.74	40.11	43.19	46.96	49.64
28	37.92	41.34	44.46	48.28	50.99
29	39.09	42.56	49.59	52.34	52.34
30	40.26	43.77	50.89	53.67	53.67
40	51.81	55.76	63.69	66.77	66.77
50	63.17	67.50	76.15	79.49	79.49
60	74.40	79.08	88.38	91.95	91.95
70	85.53	90.53	100.42	104.22	104.22
80	96.58	101.88	112.33	116.32	116.32
90	107.56	113.34	124.12	128.30	128.30
100	118.50	124.34	129.56	135.81	140.17

جدول 7/1 القيم الحرجة لاختبار مان – وتنــي على مستوى دلالة $\alpha = 0.05$ *

القيم الحرجة لاختبار أحادي الجانب على مستوى α = **0.05** تشير إليها القيم غير المظللة بينما القيم الحرجة لاختبار ثنائي الجانب تشير إليها القيم المظللة . ولكي تكون U دالة يجب أن تكون مساوية أو أقل من القيم الحرجة في الجدول . الشرطات (–) داخل الجدول تشير إلى أنه ليس هناك إمكانية اتخاذ قرار على مستوى دلالة محدد وقيم n_A و n_B .

n_B \ n_A	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0	0
	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
2	–	–	–	–	0	0	0	1	1	1	1	2	2	2	3	3	3	4	4	4
	–	–	–	–	–	–	–	0	0	0	0	1	1	1	1	1	2	2	2	2
3	–	–	0	0	1	2	2	3	3	4	5	5	6	7	7	8	9	9	10	11
	–	–	–	–	0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8
4	–	–	0	1	2	3	4	5	6	7	8	9	10	11	12	14	15	16	17	18
	–	–	–	0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	13
5	–	0	1	2	4	5	6	8	9	11	12	13	15	16	18	19	20	22	23	25
	–	–	0	1	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
6	–	0	2	3	5	7	8	10	12	14	16	17	19	21	13	15	16	28	30	32
	–	–	1	2	3	5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
7	–	0	2	4	6	8	11	13	15	17	19	21	24	26	28	30	33	35	37	39
	–	–	1	3	5	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
8	–	1	3	5	8	10	13	15	18	20	23	26	28	31	33	36	39	41	44	47
	–	0	2	4	6	8	10	13	15	17	19	22	24	26	29	31	34	36	38	41
9	–	1	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54
	–	0	2	4	7	10	12	15	17	20	23	26	28	31	34	37	39	42	45	48
10	–	1	4	7	11	14	17	20	24	27	31	34	37	41	44	48	51	55	58	62
	–	0	3	5	8	11	14	17	20	23	26	29	33	36	39	42	45	48	52	55
11	–	1	5	8	12	16	19	23	27	31	34	38	42	46	50	54	57	61	65	69
	–	0	3	6	9	13	16	19	23	26	30	33	37	40	44	47	51	55	58	62
12	–	2	5	9	13	17	21	26	30	34	38	42	47	51	55	60	64	68	72	77
	–	1	4	7	11	14	18	22	26	29	33	37	41	45	49	53	57	61	65	69
13	–	2	6	10	15	19	24	28	33	37	42	47	51	56	61	65	79	75	80	84
	–	1	4	8	12	16	20	24	28	33	37	41	45	50	54	59	63	67	72	76
14	–	2	7	11	16	21	26	31	36	41	46	51	56	60	66	71	77	82	87	92

n_B \ n_A	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	–	1	5	9	13	17	22	26	31	36	40	45	50	55	59	64	67	74	78	83
15	–	3	7	12	18	23	28	33	39	44	50	55	61	66	72	77	83	88	94	100
	–	1	5	10	14	19	24	29	34	39	44	49	54	59	64	70	75	80	85	90
16	–	3	8	14	19	25	30	36	42	48	54	60	65	71	77	83	89	95	101	107
	–	1	6	11	15	21	26	31	37	42	47	53	59	64	70	75	81	86	92	98
17	–	3	9	15	20	26	33	39	45	51	57	64	70	77	83	89	96	102	109	115
	–	2	6	11	17	22	28	34	39	45	51	57	63	67	75	81	87	93	99	105
18	–	4	9	16	22	28	35	41	48	55	61	68	75	82	88	95	102	109	116	123
	–	2	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99	106	112
19	0	4	10	17	23	30	37	44	51	58	65	72	80	87	94	101	109	116	123	130
	–	2	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	119
20	0	4	11	18	25	32	39	47	54	62	69	77	84	92	100	107	115	123	130	138
	–	2	8	13	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127

جدول 7/1 القيم الحرجة لاختبار مان – وتنــي على مستوى ألفا $\alpha = 0.01$ *

القيم الحرجة لاختبار أحادي الجانب على مستوى $\alpha = 0.01$ تشير إليها القيم غير المظللة بينما القيم الحرجة لاختبار ثنائي الجانب تشير إليها القيم المظللة . ولكي تكون U دالة يجب أن تكون مساوية أو أقل من القيم الحرجة في الجدول . الشرطات (–) داخل الجدول تشير إلى أنه ليس هناك إمكانية اتخاذ قرار على مستوى دلالة محدد وقيم nA و nB .

n_B \ n_A	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	–	–	–	–	–	–	–	–	–	–	–	–	0	0	0	0	0	0	1	1
2	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0	0
	–	–	–	–	–	–	0	0	1	1	1	2	2	2	3	3	4	4	4	5
3	–	–	–	–	–	–	–	–	0	0	0	1	1	1	2	2	2	2	3	3
	–	–	–	–	0	1	1	2	3	3	4	5	5	6	7	7	8	9	9	10
4	–	–	–	–	–	0	0	1	1	2	2	3	3	4	5	5	6	6	7	8
	–	–	–	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
5	–	–	–	–	0	1	1	2	3	4	5	6	7	7	8	9	10	11	12	13
	–	–	–	1	2	3	4	6	7	8	9	11	12	13	15	16	18	19	20	22
6	–	–	–	0	1	2	3	4	5	6	7	9	10	11	12	13	15	16	17	18
	–	0	0	1	3	4	6	7	9	11	12	14	16	17	19	21	23	24	26	28
7	–	–	–	0	1	3	4	6	7	9	10	12	13	15	16	18	19	21	22	24
	–	–	0	2	4	6	7	9	11	13	15	17	20	22	24	26	28	30	32	34
8	–	–	–	1	2	4	6	7	9	11	13	15	17	18	20	22	24	26	28	30
	–	–	1	3	5	7	9	11	14	16	18	21	23	26	28	31	33	36	38	40
9	–	–	0	1	3	5	7	9	11	13	16	18	20	22	24	27	29	31	33	36
	–	–	1	2	6	8	11	13	16	19	22	24	27	30	33	36	38	41	44	47
10	–	–	0	2	4	6	9	11	13	16	18	21	24	26	29	31	34	37	39	42
	–	–	0	2	4	6	9	11	13	16	18	21	24	26	29	31	34	37	39	53
11	–	–	0	2	5	7	10	13	16	18	21	24	27	30	33	36	39	42	45	48
	–	–	2	5	8	11	14	17	21	24	28	31	35	38	42	46	49	53	56	60
12	–	–	1	3	6	9	12	15	18	21	24	27	31	34	37	41	44	47	51	54
	–	0	2	5	9	12	16	20	23	27	31	35	39	43	47	51	55	59	63	67
13	–	–	1	3	7	10	13	17	20	24	27	31	34	38	42	45	49	53	56	60

n_B \ n_A	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	–	0	2	6	10	13	17	22	26	30	34	38	43	47	51	56	60	65	69	73
14	–	–	1	4	7	11	15	18	22	26	30	34	38	42	46	50	54	58	63	67
	–	0	3	7	11	15	19	24	28	33	37	42	47	51	56	61	66	70	75	80
15	–	–	2	5	8	12	16	20	24	29	33	37	42	46	51	55	60	64	69	73
	–	0	3	7	12	16	21	26	31	36	41	46	51	56	61	66	71	76	82	87
16	–	–	2	5	9	13	18	22	27	31	36	41	45	50	55	60	65	70	74	79
	–	0	4	8	13	18	23	28	33	38	44	49	55	60	66	71	77	82	88	93
17	–	–	2	6	10	15	19	24	29	34	39	44	49	54	60	65	70	75	81	86
	–	0	4	9	14	19	24	30	36	41	47	53	59	65	70	76	82	88	94	100
18	–	–	2	6	11	16	21	26	31	37	42	47	53	58	64	70	75	81	87	92
	–	1	4	9	15	20	26	32	38	44	50	56	63	69	75	82	88	94	101	107
19	–	0	3	7	12	17	22	28	33	39	45	51	56	63	69	74	81	87	93	99
	–	1	5	10	16	22	28	34	40	47	53	60	67	73	80	87	93	100	107	114
20	–	0	3	8	13	18	24	30	36	42	48	54	60	67	73	79	86	92	99	105

جدول 8/1 القيم الحرجة لاختبار ولكوكسن * Wilcoxon T

* لكي يكون الاختبار اختباراً دالاً يجب أن تكون درجة الاختبار مساوية أو أقل من القيمة الحرجة . (–) تعني هذه الإشارة في العمود أنه لا يوجد قرار ممكن لدلالة α و N .

	مستوى الدلالة لاختبار أحادي الجانب					مستوى الدلالة لاختبار أحادي الجانب			
	.05	.025	.01	.005		.05	.025	.01	.005
	مستوى الدلالة لاختبار ثنائي الجانب					مستوى الدلالة لاختبار ثنائي الجانب			
n	.10	.05	.02	.01	n	.10	.05	.02	.01
5	0	–	–	–	28	130	116	101	91
6	2	0	–	–	29	140	126	110	100
7	3	2	0	–	30	151	137	120	109
8	5	3	1	0	31	163	147	130	118
9	8	5	3	1	32	175	159	140	128
10	10	8	5	3	33	187	170	151	138
11	13	10	7	5	34	200	182	162	148
12	17	13	9	7	35	213	195	173	159
13	21	17	12	9	36	227	208	185	171
14	25	21	15	12	37	241	221	198	182
15	30	25	19	15	38	256	235	211	194
16	35	29	23	19	39	271	249	224	207
17	41	34	27	23	40	286	264	238	220
18	47	40	32	27	41	402	279	252	233
19	53	46	37	32	42	319	294	266	247
20	60	52	43	37	43	336	310	281	261
21	67	58	49	42	44	353	327	296	276
22	75	65	55	48	45	371	343	312	291
23	83	73	62	54	46	389	361	328	307
24	91	81	69	61	47	407	378	345	322
25	100	89	76	68	48	426	396	362	339
26	110	98	84	75	49	446	415	379	355
27	119	107	92	83	50	466	434	397	373

ملحق (2)

مسرد لبعض المفاهيم الواردة في متن المصنف

تسهيلاً على طلابنا، فقد ارتأينا أن نضمن في هذا الكتاب بعض المفاهيم الإحصائية مع تعريفاتها:

1- ***Arithmetic Mean*** وسط حسابي: الوسط الحسابي هو مجموع كل الدرجات في التوزيع مقسماً على العدد الكلي للحالات.

2- ***Asymmetric Measures of association*** مقاييس التطابق اللا متماثلة: هي تلك المقاييس التي تعتمد قيمتها على أي من المتغيرات الذي حدد بأنه متغيرٌ مستقلٌ، وأي من المتغيرات الذي حدد بأن يكون تابعاً.

3- ***Binomial distribation*** التوزيع ذو الحدين: هو ذلك التوزيع الذي يمتلك فقط قيمتين أو فئتين احتماليتين.

4- ***Bivariate descriptive Statistics*** الإحصاءات الوصفية الثنائية: هي نوع من الإحصاءات التي يمكن استخدامها في التحليل لمعرفة ما إذا كانت هناك علاقة بين متغيرين.

5- ***Bivariate table*** الجدول الثنائي: هو ذلك الجدول الذي يعرض التوزيع التكراري المشترك لمتغيرين.

6- ***Case*** حالة: هي كينونة تعرض أو تمتلك خاصيات متغير. أي خاصية تعكس عملية القياس المرتبطة بالمتغير.

7- ***Central Limit Theorm*** نظرية النهائية المركزية: هي تلك النظرية التي تبين أنه إذا كان عدد العينات العشوائية المأخوذة من مجتمع ما، وذات أحجام متساوية واللا متناهية العدد، فإن توزيع المعاينة لمتوسطات العينات يميل إلى أن يكون موزعاً – تقريباً – توزيعاً طبيعياً.

-8 **Class Interval** فئات متساوية: الفئات المتساوية هي عبـارة عـن مـدى مـن القـيم على توزيع تم تصنيف هذه القيم معاً للعرض والتحليل.

-9 **Coefficient of relative Variation (CRV)** معامل التباين النسبي: معامل التباين النسبي هو عبارة عن إحـصاءات وصـفية تعـبر عـن الانــحراف المعيـاري للتوزيـع كنسبة للمتوسط الحسابي.

-10 **Determination (r^2)** معامل التحديد (r^2): هـي نـسبة كـل التبـاين في y الـذي تم تفسيره من خلال X. ويمكننا إيجاد هذا التباين بتربيع قيمة ارتباط بيرسون (r).

-11 **Coefficient of Mnltiple determination(R^2)** معامل التحديد المتعدد: هـو ذاك الإحصاء المساوي للتباين الكلي الذي تم تفـسيره في المـتغير التـابع مـن خـلال كـل المتغيرات المستقلة مجتمعة.

-12 **Colum** عمود: العمـود هـو البعـد العمـودي للجـدول الثنـائي. وحـسب العـرف الأكاديمي يمثل كل عمود درجة على المتغير المستقل.

-13 **Conceptual definitions** التعريفـات التـصورية: التعريفـات التـصورية هـي تلـك التعريفات التي تحدد خاصيات متغير (يطلق عليها أيضاً التعريف الاسمي).

-14 **Concordant Pair** الأزواج المتوافقة: تعني الأزواج المتوافقـة أن حـالتين في توزيـع مشترك تم ترتيبهما على كلا المتغيرين بشكل متساوٍ.

-15 **Confidence Level** مستوى الثقة: يعني مستوى الثقة الاحتماليـة بـأن تقـدير فـترة سوف يتضمن قيمة معلمة المجتمع التي تم تقديرها.

-16 **Constant** ثابت: الثابت هو عبارة عن الخاصية التي لا تتبـاين مـن حالـة إلى حالـة أخرى.

-17 **Contingency Table** جدول التوافق: (انظر تعريف الجدول الثنائي في هذا المسرد).

-18 **Continous Variable** المتغير المتصل: المتغير المتصل هو ذلـك المـتغير الـذي يأخـذ جميع القيم بين حدي المتغير، وإذا زاد من رقم إلى الرقم التالي له يمـر بكـل الكـسور الممكنة بينهما (العمر على سبيل المثال).

19- ***Coordinate*** متساوٍ: هي نقطة على شكل الانتشار تـشير بـشكل تلقـائي إلى القـيم لحالة معينة تدون لكل متغير.

20- ***Correlation Matrix*** مصفوفة الارتباط: عبارة عن جدول يـبين الارتباطـات بـين كل المتغيرات الممكنة المؤتلفة.

21- ***Critical Region*** المنطقة الحرجة: تعني المنطقة الحرجة مدى الـدرجات الـذي تَـسَبَّبَ في رفض الفرض الصفري (Ho) على مستوى محدد من الدلالة (α).

22- ***Crosstabulation*** التقاطع: (انظر تعريف الجدول الثنائي في هذا المسرد).

23- ***Cumulative Frequency Table*** الجدول التكراري التجمعي (التراكمـي): يـشير الجدول التكراري التجمعي إلى كل قيمة في التوزيع وتجمع أعداد الفئات إلى بعـضها البعض لتشمل تلك القيمة.

24- ***Cumulative relative Frequency Table*** الجدول التكراري المجمع النسبي: هي تلك الجداول التي تعبر عن الحالات داخـل كـل قيمـة مـن قـيم المـتغير كنـسبة مئويـة أو تناسب من المجموع الكلي للحالات تضاف إلى بعضها البعض لتشمل تلك القيمة..

25- ***Data*** بيانات: هي تلك المعلومات التي يقوم الباحث بجمعهـا مـن خـلال قـضية أو مشكلة يطرحها للبحث والاستقصاء.

26- ***Degree of Freedom*** درجـة الحريـة: مفهـوم إحـصائي يـشير إلى عـدد القـيم في التوزيـع الـذي يكـون حـراً في تباينـه فيمـا يتعلـق بمتوسـط عينـة، توزيـع لـه N-1 درجة حرية، هذا يعني أنه فيما يتعلـق بقيمـة محـددة المتوسـط. N-1 درجـات حرة في تباينها. على سبيل المثال، لو كان لدينا متوسط يساوي 3 و 5 =N. فالتوزيع لخمس درجات سيكون لديها N-1 أو 4 درجات حرية. عندما تكون قيم الدرجات 1، 2، 3، 4، فإن الخامسة يجب أن تكون 5 ولا قيم أخرى.

27- ***Dependent Sample*** عينات تابعة: العينات التابعـة هـي تلـك العينـات الـتي يـتم تحديدها من خلال تركيبة عينات أخرى.

28- **Descriptive Statistics** إحــصاءات وصــفية: تــشير الإحــصاءات الوصــفية إلى الأعداد والتمثيل البياني، وتقنيات الجدولة لتنظيم البيانات، وعرضها، وتحليلها.

29- **Dichotomous Variable** المتغير الثنــائي: هــو ذلــك المـتغير الـذي يمتلـك قيمـتين احتماليتين فقط.

30- **Directional Hypothesis** الفرضية الاتجاهية: هي تلك الفرضية التي يتوقع الباحـث من خلالها وجود علاقة محددة بين متغيرين (سالبة أو موجبة).

31- **Disconcordant Pair** الأزواج المتنافرة: تعني الأزواج المتنافرة أن حالتين في توزيع مشترك تم ترتيبهما على متغير بشكل مختلف عن بعضهما البعض.

32- **Discrete Variable** المتغير المنفصل: هو ذلك المتغير الذي لا تُمَكِّنُهُ طبيعته مـن أن يأخذ جميع القيم بـين حـدي المـتغير، بـل تتحـرك عـن أعـداد معينـة دون سـواها ؛ وبالتالي فإن المتغير المنفصل لا يحتوي على كسور (عـدد الأطفـال في الأسـرة علـى سبيل المثال).

33- **Dummy Variable** المتغيرات الديميوية: هي متغيرات مقاسة على أي مـستوى مـن القياس بما فيها المقياس الاسمي. ولديها فئتان يرمز للفئة الأولى بـصفر (0)، والفئـة الثانية بِـ (1) وذلك لتضمينها في معادلة الانــحدار.

34- **Effect Size** حجم التأثير: عبارة عن مجموعـة مـن الإحـصاءات الـتي تحـدد القـوة النسبية للفروق بين قيم الوسط الحسابي. بمعنى آخر، يعني حجم التأثير مقدار التباين الكلي في المتغير التابع الذي يمكن التنبؤ به من معرفة مستويات المتغير المستقل.

35- $\boldsymbol{Eta^2 = n^2}$ إيتا تربيع: يمثل إحصاء Eta^2 نسبة التباين في المـتغير التـابع الـتي يفـسرها ويوضحها المتغير المستقل، وتتراوح قيمتها من 0 إلى 1.0.

36- **Frequency** التكرار: هو عدد المرات التي تظهر فيها درجة مفردة في مجموعة بيانات.

37- **Frequency Table** الجدول التكراري: هي تلـك الجـداول الـتي تلخـص إجابـات المبحوثين في كل فئة من فئات المتغير المدروس.

-38 **Gamma** جاما (G): تعتبر جاما مقياساً للتطـابق للمـتغيرات الترتيبيـة (والثنائيـة). ويشار إلى جاما (G) في بعض الأحيان بـِ Goodman-Kruskalś Gamma كمقيـاس متماثل للتطابق. وتتراوح قيمة جاما من 1 – إلى 1.0+. وتشير قيمة 1+ إلى تطـابق تـام، حيث إن كل التباين في المتغير التابع يعزى إلى التباين في المتغير المستقل. أما قيمة 1– فهي تشير، مرة أخرى، إلى وجود تطابق حتمي، ولكنه تطـابق سـالب. بينمـا كـل التبـاين في المتغير التابع يعزى للتباين في المتغير المستقل، فالتطابق يتجه في الاتجاه المعاكس (سالب). وعندما تشير قيمة جاما إلى صـفر (0) يعـني لا يوجـد تطـابق، لا شـيء في التبـاين في المتغير التابع يمكن أن يفسر من خلال التباين في المتغير المستقل.

-39 **Goodness of fit test** اختبار جودة التطـابق: يبحـث هـذا الاختبـار مـا إذا كانـت بيانات عينة ما (fo) تتطابق مع توقعات معينة (fe) يُفترض أن يتميز بها مجتمع ما.

-40 **Hypothesis** الفرض: هو عبارة عن بيان حول بعض خصائص التوزيـع في مجتمـع إحصائي.

-41 **Hypothesis testing** اختبار الفرض: هو الإجـراء الـذي يحـدد مـن خلالـه مـا إذا كانت بعض مظاهر توزيع مجتمع إحصائي تمتلك خاصيات محددة.

-42 **Independence** الاسـتقلال: يعتـبر المـتغيران مـستقلين إذا كـان نمـط التبـاين في الدرجات غير مرتبط بنمط التباين في الدرجات للمتغير الآخر.

-43 **Index of qualitative Variation (IQV)** مؤشـر التبـاين النـسبي: هـو نـوع مـن الإحصاء يخبرنا كم من المتغيّرية توجد في متغير لديه فئات اسمية وأن مدى هذا المقيـاس (IQV) تمتد قيمته من 0 – 1.0. وكلما كانت القيمة قريبة من صـفر، كانـت هنـاك أقـل متغيرية في توزيع المبحوثين على المتغير، وكلما قربت القيمة من 1.0، زادت المتغيّرية.

-44 **Inferential Statistics** الإحـصاء الاسـتنتاجي (الاسـتدلالي): يـشير الإحـصاء الاستنتاجي (الاستدلالي) إلى التقنيات العدديـة الـتي تـستخدم للوصـول إلى نتـائج حول توزيع مجتمع إحصائي استناداً على بيانات من عينة عـشوائية، تم سـحبها مـن ذلك المجتمع الإحصائي.

-45 **_Interqnartile range_** المدى الربيعي: هو الفرق بين الحـدود العليـا للربيـع الأول، والربيع الثالث ؛ أي المدى الوسط 50 % للحالات في مجموعة مرتبـة مـن التوزيـع ($R = Q_3 - Q_1$).

-46 **_Interval Scale_** المقياس ذو المسافات المتساوية: هو المقياس الـذي يمتلـك وحـدات تم قياسها على ذي المسافات المتساوية ولديها مسافة متساوية بين القيم في المقياس.

-47 **_Least Significance difference (LSD)_** أقل فرق الدلالة: هي قيمة عددية يمكن حسابها إذا كان التكرار داخل كل معاملة متساوياً:

$$LDS = t\sqrt{\frac{2\,s^2}{r}}$$

أما إذا كانت التكرارات داخل كل معاملة غير متساوية:

$$LDS = t\sqrt{\frac{S^2}{r_1}+\frac{S^2}{r_1}}$$

-48 **_Lambda_** لامبيـدا: هـي مقيـاس للتطـابق لنـسبة التخفـيض في الخطـأ (PRE) للمتغيرات الاسمية. وتتراوح قيمتها من 0 إلى 1.0. وعندما تكـون لامبيـدا مـساوية لصفر (0) يعني أنه لا يوجد تطابق ؛ أي لا شيء في التباين في المتغير التابع يمكـن أن يفسر من خلال التباين في المتغير المستقل. أما إذا كانت قيمة لامبيدا مساوية لِـ 1.0، فهذا يعني وجود تطابق تام 100 % أي أن كل التباين في المتغير التابع يمكـن تفـسيره بواسطة التباين في المتغير المستقل.

-49 **_Least Square Method_** طريقة أقل التربيعات: طريقة تـستخدم في توضـيح الخـط المستقيم الأمثل الذي يمثل العلاقة بين المتغير X والمتغير y.

-50 **_Level of Significance_** مستوى الدلالة: هو أقصى احتمال الذي يحتمل بمقتـضاه مخاطرة الوقوع في الخطأ من النوع الأول. أي القيمة المحددة التي علـى ضـوئها يُقبـل أو يُرفض الفرض الصفري (Ho).

51- ***Measurment*** القياس: هـو العمليـة الـتي مـن خلالهـا يـستطيع الباحـث أن يحـدد ويسجل الخاصيات الممكنة للمتغير كحالة مفـردة. بمعنـى آخـر، يعـني القيـاس بأنـه طريقة خاصة تتبع في قيـاس المـتغيرات والمفـاهيم الاجتماعيـة (مـتغير النـوع يأخـذ قيمتين ذكراً أو أنثى، ويحدد لنا القياس ما هي الفئة التي يقع فيها هذا الشخص).

52- ***Measurment of association*** مقياس التطابق: إحصاءات وصفية تشير إلى مـدى التغير في قيم متغير مرتبطة بالتغير في قيم متغير آخر.

53- ***Measures of Central Tendency*** مقـاييس النـــزعة المركزيـة: هـي تلـك الإحصاءات الوصفية التي تشير إلى قيمة المعدل في التوزيع.

54- ***Measures of dispersion*** مقاييس التشتت: هي تلك الإحصاءات الوصـفية الـتي تشير إلى تباين الدرجات في التوزيع.

55- ***Median*** الوسيط: هـو أحـد مقـاييس النـــزعة المركزيـة الـذي يـشير إلى القيمـة في مجموعة مرتبة، وتقسم التوزيع إلى نصفين.

56- ***Mode*** المنوال: هو القيمة الأعلى تكراراً في التوزيع. أي القيمة التي تتكرر أكثر مـن غيرها.

57- ***Multiple Correlation*** الارتباط المتعدد: عبارة عن تقنيات إحصائية تشير إلى قـوة بين متغير تابع واثنين أو أكثر من المتغيرات المستقلة.

58- ***Multivariate Regression*** الانـحدار المتعدد: هو تلـك التقنيـات الـتي تبحـث في العلاقة بين متغيرين مستقلين أو أكثر وعلاقتها بمتغير تابع واحد.

59- ***Nominal definition*** التعريف الاسمي: (انظر تعريف التعريف التصوري في هـذا المسرد).

60- ***Nominal Scale*** المقياس الاسمي: مجموعة فئـات لـديها أسمـاء مختلفـة والقيـاس على المستوى الاسمي يصف ويصنف المشاهدات، ولكنه لا يحدد أي تمييز كمي بين هذه المشاهدات.

61- *Non-directional Hypothesis* الفرضية غير الاتجاهية: هي تلك الفرضية الـتي لا يتوقع الباحث من خلالها أي علاقة محددة بين متغيرين خاضعين للدراسة.

62- *Nonparametric test* الاختبار اللا معلمي (اللا بـارامتري): هـو اختبـار لغـرض حول خصائص توزيع مجتمع أكثر منه كاختبار معلمات مجتمع.

63- *Normal distribution* التوزيع الطبيعي: هو التوزيع النظري للـدرجات المتماثلـة، أحادية النموذج، وذات الشكل الجرسي، والمنحنى الطبيعـي المعيـاري دائمـاً يكـون متوسطه صفراً، وانـحرافه المعياري واحد (1).

64- *Operational definition* التعريف الإجرائي: هو مجموعة الإجراءات الـتي تـصف النشاطات التي يجب القيام بها من أجل تعيين الأبعاد التي يمكن قياسـها وملاحظتهـا من أجل التعرف على مـا يـشير إليـه هـذا المفهـوم أو ذاك في الدراسـة الـتي يجريهـا الباحث.

65- *Ordinal Scale* مقياس تـرتيبي: هـو أحـد مـستويات القيـاس ويـشير، إضـافة إلى خاصية التصنيف، أنه يسمح للحالات أن ترتب وفقاً لدرجة كل فئة، وطبقاً لقيـاس المتغير محل اهتمام الباحث.

66- *Ordinary Least Square Regression*: عبارة عن قاعدة تبين أفضل خط مطـابق للانـحدار الخطي، ذلك الخط الذي يقلل مجموع تربيعات المتبقي.

67- *Parameter* معلمة: المعلمة عبارة عن إحصاء يصف بعض معالم المجتمع: المتوسـط، الانـحراف، التباين... الخ. (متوسط الدخل، والتوزيع العمري لمدينة ما).

68- *Parametric Test* اختبار مَعْلمي (بـارامتري): الاختبـار المعلمـي هـو عبـارة عـن اختبار لفرض حول معلمات توزيع مجتمع إحصائي.

69- *Partial Correlation* الارتباط الجزئي: هو تلك المعامل التي يستطيع الباحث مـن خلالها تحييد تأثير المتغير الثالث عندما يبحث في العلاقة بين متغيرين X و y.

70- **_Partial Correlation Coefficient_** معامل الارتباط الجزئي: معامل الارتباط الجزئـي هو إحصاء يبين العلاقة بين مـتغيرين، بينمـا الـتحكم في مـتغيرات أخـرى. والرمـز $r_{yx.A}$ يشير لمعامل الارتباط الجزئي عند التحكم في متغير واحد.

71- **_Percentage_** النسب: النسبة عبارة عن إحصاء يعاير الحجم الكلي للحـالات علـى أساس قيمة مائة (100).

72- **_Perfect Association_** تطابق تام: التطابق التام يعبر عن علاقة إحـصائية حيـث إن كل الحالات التي لديها قيم مفردة لمتغير واحد يمتلك قيمة محددة للمتغير الآخر.

73- **_Population_** مجتمع إحصائي: هو مجموع الحالات التي يرغب الباحث في دراستها.

74- **_Post-hoc Tests_**: هي اختبارات إضافية التي تجرى بعد اختبـار ANOVA لتحديـد أي من فروق المتوسط تكون دالة وأي منها لا تكون كذلك.

75- **_Proportional Reduction in Error_** نسبة التخفيض في الخطـأ: نـسبة التخفـيض في الخطأ هو المنطق الذي يشكل التعريـف وحـساب لامبيـدا وجامـا. وهـو الإحـصاء الذي يقارن عدد الأخطاء التي حدثت عند التنبؤ بالمتغير التـابع عنـد تجاهـل المـتغير المستقل، مع عدد الأخطاء التي حدثت عند الأخذ في الاعتبار المتغير المستقل.

76- **_Proportions_** التناسب: عبارة عن إحصاء لمعـايرة الحجـم الكلـي للحـالات علـى أساس قيمة واحد (1).

77- **_Random Selection_** اختبار عشوائي: هي عملية المعاينـة حيـث إن كـل مفـردة في المجتمع يكون لديها نفس فرصة الاختيار في العينة.

78- **_Range_** المدى: هو الفرق بين أكبر قيمة وأصغر قيمة في التوزيع.

79- **_Rank_** ترتيب: هو العدد الذي يشير إلى وضع حالة في سلسلة مرتبة.

80- **_Ratio Scale_** المقياس النسبي: هو أحد مستويات القياس الـذي يعطـي قيمـة صـفر (0) للحالات التي تظهر لا كمية في المتغير.

-81 ***Region of Rejection*** منطقة الرفض (انظر Critical region من هذا المسرد).

-82 ***Regression Coefficient*** معامــل الانــحدار: معامــل الانــحدار عبــارة عــن إحصاءات وصفية تشير إلى كم من الوحدات في المتغير التابع ستتغير مع تغير وحـدة معطاة في المتغير التابع.

-83 ***Regression Line*** خط الانـحدار: هو أفضل خط تطابق مستقيم يلخـص العلاقـة بين متغيرين. فخطوط الانـحدار مطابقة نقاط البيانات عـبر مقيـاس أقـل المربعـات الذي به يلمس فيه الخط كل متوسط شرطي لِـ: y. أو تكون قريبة من الخط إلى حـد ممكن لكل الدرجات. (ويصلح خط الانـحدار في عمليـة التنبـؤ مـن خـلال معرفـة قيمة أحد المتغيرين، فإنه بالتالي يمكن التنبؤ بقيم المتغير الثـاني. ويمكـن رسـم خـط الانـحدار عن طريق معامل الانـحدار).

-84 ***Relative Frequencies*** التوزيعات النسبية: هي إحصاءات تبين عدد الحـالات داخـل كل قيمة متعلقة بمتغير ما كنسبة أو تناسب من العدد الكلي للحالات.

-85 ***Research Hypothesis*** الفـرض البحثـي: الفـرض البحثـي عبـارة عـن صـياغة الفرض الصفري (Ho) وفي سياق دلالة الاختبـارات لعينـة مفـردة، يـشير الفـرض البحثي إلى أن المجتمع الذي سحبت منه العينة لا يمتلك خاصية أو قيمة محددة.

-86 ***Residual*** متبقٍ: المتبقي هو الفرق بين القيمة المـشاهدة والقيمـة المتوقعـة لمـتغير مـا خاضع للدراسة.

-87 ***Sample*** عينة: العينة هي مجموعة الحـالات الـتي لا تحتـوي علـى كـل الحـالات في المجتمع الإحصائي، وإنما جزء من هذا المجتمع.

-88 ***Sampling Distribution*** توزيع المعاينة: هـي توزيعـات تكراريـة يمكـن الحـصول عليها من حساب المتوسطات لكل العينات النظريـة المحتملـة للحجـم المعـين الـذي أمكن سحبه من مجتمع إحصائي محدد.

89- **Sampling Error** خطأ المعاينة: يعني خطأ المعاينة درجة الخطأ المتوقعة في احتماليـة المعاينة لعينة محددة التصميم. والمعادلة التي تحـدد خطـأ المعاينـة تحتـوي علـى ثلاثـة عوامل: المعلمة، حجم العينة، والخطأ المعياري.

90- **Scatter Plot** رسم انتشاري: هو عبارة عن تقنية بيانية لوصف توزيع مشترك لعدد متغيرين متصلين.

91- **Somerś d**: هو عبارة مقيـاس توزيعـات غـير متماثلـة التطبيـق يأخـذ في الحـسبان الرتب المشتركة علـى المـتغير y دون المـتغير X إلى جانـب الحـسابات المـستخدمة في معامل GAMMA. يضاف T_y إلى مقام معادلة جاما. وتشير T_y إلى الرتب المتعادلـة على المتغير y دون المتغير X.

ولما كان مقياس سومرز d يأخـذ في اعتبـاره الرتـب المتعادلـة علـى y دون X، فقـد حدث تناقض شديد في درجة الارتباط عند مقارنته بجاما.

92- **Spurious Relationship** علاقة كاذبة: العلاقـة الكاذبـة هـي علاقـة متعـددة بـين متغيرات حيث لا توجد بينها علاقة سببية حقيقية بين المتغيرات المستقلة والمـتغيرات التابعة. فالعلاقة بين هذه المتغيرات كانت نتيجة لبعض المتغيرات الأخرى.

93- **Standard Deviation** انـحراف معياري: الانـحراف المعيـاري هـو أحـد مقـاييس التشتت الذي يشير إلى الجذر التربيعي للتباين.

94- **Standard Error** الخطـأ المعيـاري: الخطـأ المعيـاري هـو الجـذر التربيعـي لمتوسـط مجموع مربعات انـحراف القيم عن متوسطها الحسابي.

95- **Standardizing Distribution** معايرة التوزيع: تعني معايرة التوزيع ذلك التوزيـع الطبيعي الذي يتم فيه تحويل الدرجات الإحـصائية إلى درجـات معياريـة تعـبر عـن درجـات أي توزيـع تكـراري طبيعـي. وتـسمح إجـراءات المعـايرة بقيـاس كـل التوزيعات الطبيعيـة في إطـار الوحـدات الـشائعة، وحـدات الانــحراف المعيـاري، بغض النظر عن الوحدات التي يتم قياسها.

96- **_Statistics_** الإحصاء: عبارة عن تقنيات رياضية تستخدم لفحص البيانات من أجل الإجابة على السؤال المطروح أو اختبار نظريات.

97- **_Symmetric Measures of association_** مقاييس التطابق المتماثل: مقاييس التطابق المتماثل هي تلك المقاييس التي تكون قوتها متساوية بغض النظر عن أي من المتغيرين قد حدد كمتغير مستقل أو أي منهما قد حدد كمتغير تابع.

98- **_Tabulation_** جدولة: الجدولة هي العملية التي بمقتضاها يستطيع الباحث وضع كل حالة من الحالات المدروسة في المكان المخصص لها في التوزيع التصنيفي.

99- **_Test of Independence_** اختبار الاستقلال: اختبار الاستقلال هو دراسة العلاقة أو العلاقات بين ظاهرتين أو أكثر، وبيان ما إذا كان توزيع هاتين الظاهرتين أنهما مستقلتان أو غير مستقلتين عن بعضهما البعض.

100- **_Test Statistic_** إحصائي الاختبار: إحصائي الاختبار هو قيمة إحصائية تسحب من بيانات العينة المدروسة، وتستخدم كأداة للوصول إلى قرار قبول أو رفض فرضية إحصائية معطاة، وذلك اعتماداً على القيمة الإحصائية المسحوبة. وتأخذ هذه القيمة الإحصائية عديداً من القيم الاحتمالية اعتماداً على طبيعة العينة العشوائية المستخدمة في الدراسة.

101- **_Total Variation_** التباين الكلي: هو انتشار درجات y حول متوسط y.

102- **_Type I Error or (Alph Error)_** الخطأ من النوع الأول (خطأ ألفا): يعني الخطأ من النوع الأول رفض الفرض الصفري "لا فرق" عندما يكون في الواقع فرضاً صحيحاً.

103- **_Type II Error or (Beta Error)_** الخطأ من النوع الثاني أو خطأ بيتا: يعني الخطأ من النوع الثاني أن الباحث ليس في مقدوره رفض الفرض الصفري عندما يكون في الواقع فرضاً خطأ.

-104 **_Variable_** متغير: المتغير هو عبارة عن خاصية تجريبية تتخذ قيمتين أو أكثر. فإذا كانت هذه الخاصية قابلة للتغير كماً أو نوعاً ينظر إليها كمتغير (العمر، الدخل).

-105 **_Variability_** المتغيّرية: المتغيرية تمدنا بقياس كمي لمعرفة إلى أي مدى تكون الدرجات في التوزيع قد انتشرت أو تجمعت معاً، ومن هنا تعرف المتغيّرية من خلال المسافة، حيث يمكننا معرفة كم المسافة التي نتوقعها بين درجة وأخرى. وكم المسافة المتوقعة بين درجة الفرد والمتوسط الحسابي. ويعتبر مفهوم المتغيّرية مفهوماً مهماً في الإحصاءات الاستدلالية. وتحتوي المتغيرية على المدى، المدى الربيعي، والانحراف المعياري. بمعنى آخر، إذا كانت الدرجات في التوزيع كلها متساوية حينئذٍ لا توجد متغيّرية. أما إذا تمت ملاحظة فروق بسيطة بين الدرجات عندئذٍ تكون المتغيّرية قليلة. وأما إذا تمت ملاحظة فروق كبيرة بين الدرجات حينها تكون المتغيّرية كبيرة.

-106 **_Variance_** تباين: التباين هو الإحصاء الذي يعبر عن متوسط انحراف الدرجات عن المتوسط الحسابي للتوزيع.

-107 **_y - Intercept(a)_** : هي النقطة التي يمر بها خط الانحدار لمحور y.

-108 **_Z - Score_** درجة **_Z_**: درجة Z هي الدرجات المعيارية التي تعبر عن المسافة بين نقطة. ومتوسط توزيع طبيعي كنسبة للانحراف المعياري من ذاك التوزيع الطبيعي والنظري.

ملحق (3)
منظم إحصائي

الملحق التالي يبين لنا ملخصاً منظماً للأساليب الإحصائية التي تم تغطيتها في هذا الكتاب. وقد قُسم هذا المنظّم الإحصائي إلى أربعة أقسام. وكل قسم من هذه الأقسام يوضح الأساليب الإحصائية التي تخدم غاية مشتركة. والأقسام الأربعة هي

1- الإحصاء الوصفي.
2- الاختبارات البارامترية لفروق المتوسط.
3- الاختبارات اللابارامترية للفروق النظّامية.
4- تقييم العلاقة بين المتغيرات.

أولاً: الإحصاء الوصفي:

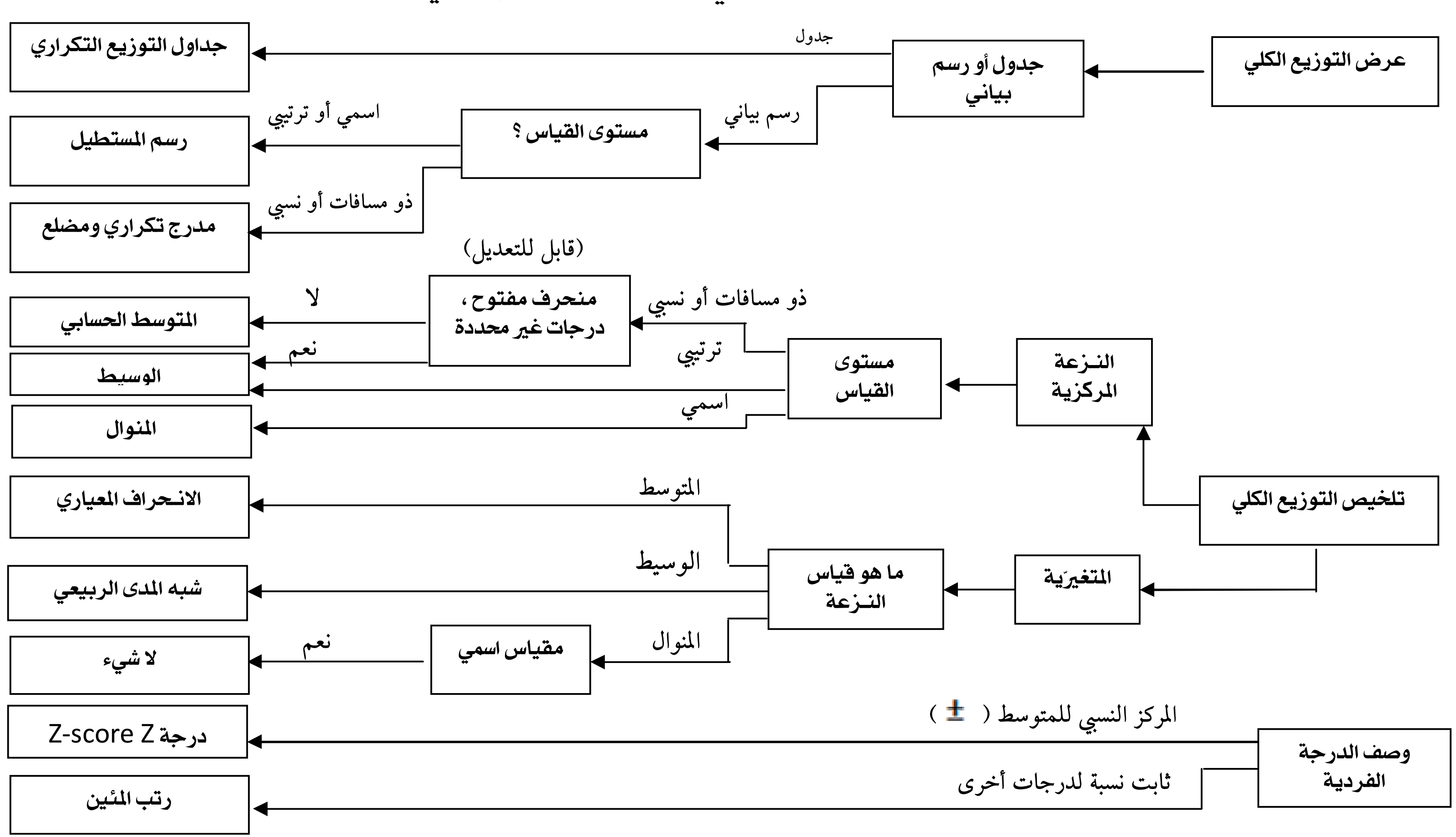
إن الهدف من وراء استخدام الأساليب الإحصائية الوصفية هو تسهيل وتنظيم مجموعة الدرجات، فالدرجات يمكن تنظيمها في جداول إحصائية أو رسوم بيانية، أو يمكن تلخيصها من خلال حساب قيمة أو قيمتين تصفان مجموعة البيانات بكاملها، وأن أكثر الأساليب الإحصائية شيوعاً واستخداماً تتمثل في(*):

أ - جداول التوزيعات التكرارية أو الرسوم البيانية.
ب- مقاييس النـزعة المركزية.
ج- مقاييس التشتت (المتغيرّية).
د- درجات Z-Scores Z.

والشكل (1) خارطة قرار تساعد الباحث في اختيار الأساليب الإحصائية الوصفية لتحليل بياناته.

(*) يمكنك الرجوع إلى المسرد الذي يتضمن تعريفات هذه المفاهيم الإحصائية المرتبطة بالإحصاء الوصفي ملحق (2).

شكل (3-1): خارطة قرار تساعد الباحث في اختيار الإحصاء الوصفي لتحليل بياناته



ثانياً: الاختبارات البارامترية: تقييم فروق المتوسط بين مجتمعين:

كل الاختبارات التي يغطيها هذا الجزء تستخدم المتوسط الذي يتم الحصول عليه من بيانات عينة كأساس لاختبار الفروض حول متوسط المجتمع. وبالرغم من وجود عدة مواقف بحثية، إلا أنها في ذات الوقت تستخدم المنطق الأساسي نفسه، وأن كل الاختبارات الإحصائية لديها نفس البناء الأساسي، وفي كل حالة، فإن إحصاء الاختبار (Z، t أو F) تشتمل على حساب نسبة ratio تبعاً للبناء التالي:

$$\text{إحصاء الاختبار} = \frac{\text{الفرق المتحصل عليه بين المتوسطات}}{\text{فرق المتوسط المتوقع بالصدفة}}$$

إن الهدف من وراء كل اختبار هو تحديد ما إذا كانت فروق المتوسط المشاهد تكون أكبر من تلك الفروق بالصدفة. وبعبارات عامة، فإن نتيجة الدلالة تعني أن النتائج المتحصل عليها في دراسة بحثية (فروق المتوسط) هي أكثر مما يتوقع من الفروق التي جاءت عن طريق الصدفة. وفي كل حالة فإن القيمة الكبيرة لنسبة إحصاء الاختبار تشير إلى نتيجة دالة.

إن الحسابات الفعلية تختلف قليلاً عن إحصاء اختبار لآخر، إلا أن هذه الاختبارات في المحصلة النهائية تشتمل كلها على نفس الحسابات الأساسية.

1- مجموعة درجات (عينة) تم الحصول عليها من كل مجتمع أو من كل حالة معالجة.

2- يتم حساب المتوسط الحسابي لكل مجموعة درجات، وأن بعض مقاييس المتغيّرية SS (مجموع التربيعات، الانـحراف المعياري أو التباين) تم الحصول عليه لكل مجموعة فردية من الدرجات.

3- الفروق بين متوسط العينات يمدنا بقياس كم يوجد من فرق بين أحوال المعالجة. إن سبب فروق المتوسط هذه، يمكن أن تكون حالات المعالجة سبباً في هذه الفروق، وعادة ما يُطْلَق عليها "فروقاً نظاميةً أو فروقاً متنبأ بها". هذه الفروق هي بسط Numerator معادلة إحصاء الاختبار.

4- المتغيرّية داخل كل مجموعة من الدرجات تمدنا بقياس غير نظامي أو فروق غير متنبأ بها من الفروق راجعة بالأساس للصدفة. ولأن الأفراد داخل كل حالة معالجة قد تمت معالجته بشكل متساوٍ تماماً، فإنه لا يوجد أي شيء يمكن أن يكون سبباً في أن تكون درجاتهم مختلفة. وعليه، فإن أية فروقٍ مشاهدة (المتغيرّية) داخل المعالجات افترض بأنها راجعة إلى الخطأ أو الصدفة.

واضعين كل ذلك في الاعتبار، فإن نسب كل إحصاء اختبار يمكن وصفها كالتالي:

$$\text{إحصاء الاختبار} = \frac{\text{الفروق(المتغيرية) بين المعالجات}}{\text{الفروق(المتغيرية) داخل المعالجات}}$$

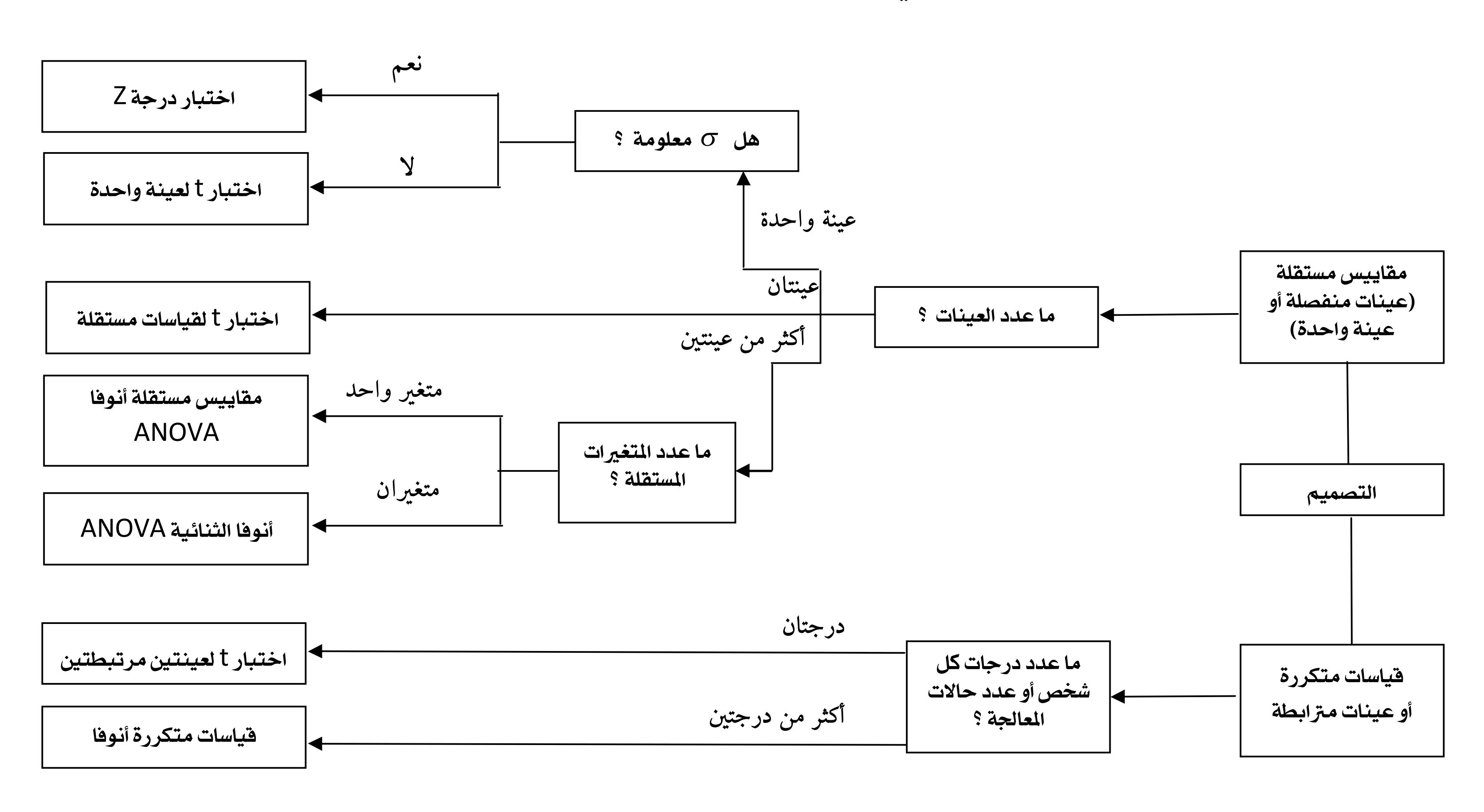
إن اختبارات الفروض التي تمت معاينتها في هذا الجزء ينطبق على ثلاثة تصميمات أساسية:

1- تصميمات لعينة مفردة (واحدة)، بيانات من عينة واحدة تستخدم لاختبار فرض حول مجتمع مفرد.

2- تصميمات مقاييس مستقلة: عينة مستقلة يتم الحصول عليها لتمثل كل مفردة أو حالة معالجة.

3- تصميمات لعينات مترابطة: إن أكثر التصميمات شيوعاً يعرف بـِ: تصميمات مقاييس متكررة التي تكون فيها نفس مجموعة الأشخاص يشاركون في كل حالات المعالجة المختلفة. وكذلك تتضمن هذه التصميمات، تصميمات الموضوعات المرتبطة التي تستخدم مجموعات منفصلة لكل حالة معالجة يتطلب فيها أن كل شخص في مجموعة واحدة يكون مرتبطاً بشكل متساوٍ مع الشخص الآخر في المجموعات الأخرى. والشكل (2) خارطة قرار تساعد الباحث في اختيار الاختبار البارامتري المناسب لتقييم فروق المتوسط بين مجتمعين.

شكل (2-3): خارطة قرار تساعد الباحث في اختيار الاختبار البارامتري المناسب لتقييم فروق المتوسط بين مجتمعين



وأخيراً، ينبغي علينا أن نكون حذرين بأن كل الاختبارات البارامترية تضع قيوداً صارمة على بيانات العينة وتوزيعات المجتمع الخاضع للدراسة. وأن أول هذه القيود، أن هذه الاختبارات تتطلب أن يكون مستوى القياس ذي المسافات أو النسبي (قيم عددية تسمح للباحث أن يقوم بحساب المتوسط، أو الفروق). وثاني هذه القيود تتمثل في أن يكون هناك افتراض لكل اختبار حول توزيع المجتمع وتقنيات المعاينة.

1- درجة Z Z- Score.
2- اختبار t لعينة واحدة.
3- اختبار t لمقاييس مستقلة.
4- اختبار t لعينتين مرتبطتين.
5- مقاييس مستقلة لأنوفا (أحادي الجانب).
6- مقاييس مستقلة لأنوفا (ثنائي الجانب).

ثالثاً: الاختبارات اللابارامترية: تقييم الفروق النظامية بين مجتمعين:

بالرغم من أن الاختبارات البارامترية هي أكثر شيوعاً واستخداماً للتقنيات الاستدلالية، إلا أن هناك عدة مواقف بحثية قد لا يكون بإمكان الباحث استخدام هذه الاختبارات البارامترية. وتندرج هذه المواقف ضمن ثلاث فئات عامة:

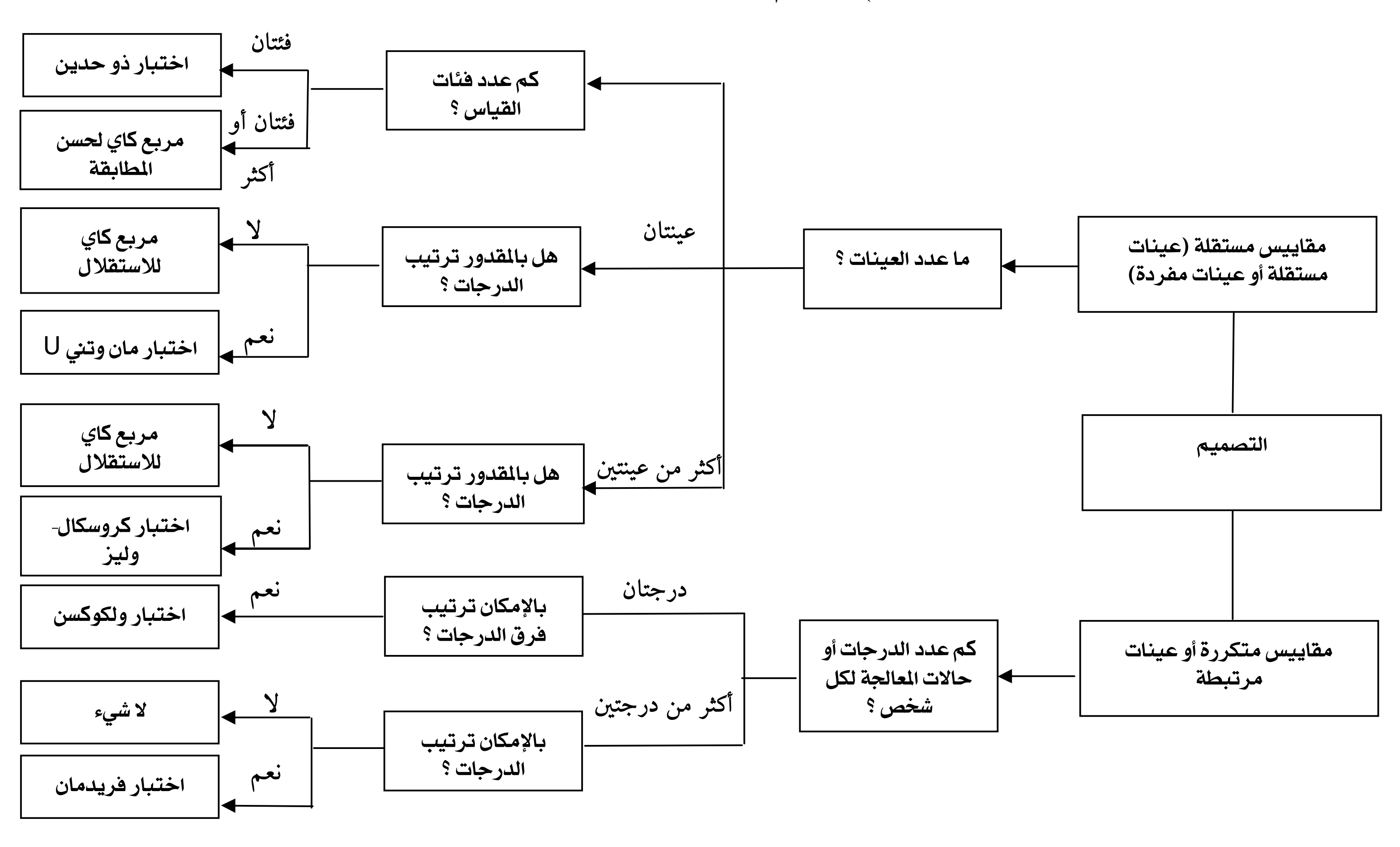
1- تتطلب البيانات مقاييس على المستوى الاسمي أو الترتيبي. وفي هذه المواقف يكون بمقدور الباحث حساب المتوسطات والتباين بإعتبار أن هذه القيم تمثل جزءاً أساسياً للاختبارات البارامترية.

2- عندما لا تَفِّي بيانات هذه الاختبارات بالافتراضات التي تستند عليها الاختبارات البارامترية.

3- تنطوي البيانات على درجة قصوى من التباين التي تفوض احتمالية الدلالة للاختبار البارامتري. في هذه الحالة، ينبغي أن تحول الدرجات إلى فئات أو رتب، وبالتالي يكون الاختبار اللابارامتري هو البديل.

وعندما يكون الاختبار البارامتري لا يمكن استخدامه، وبالتالي تكون الاختبارات اللابارامترية هي البديل الملائم. وبشكل عام، فإن الاختبارات البارامترية تعتبر اختبارات أقوى بكثير إذا ما قورنت بالاختبارات اللابارامترية.

شكل (3-3): خارطة قرار تساعد الباحث في اختيار الاختبار اللابارامتري باستخدام الدرجات غير العددية (الرتب أو فئات اسمية) لتقييم فروق نظامية بين مجتمعين



ومهما يكن الأمر، فإنه عندما لا تكـون الاختبـارات البارامتريـة غـير ملائمـة، فـإن الاختبارات اللابارامترية تمد الباحثين بتقنيـة إحـصائية بديلـة لإجـراء التحليـل والتفـسير الإحصائي للنتائج البحثية. وتشتمل الإحصاءات اللابارامترية على:

1- اختبار مربع كاي لحسن المطابقة.
2- اختبار مربع كاي للاستقلال.
3- الاختبار ذي الحدين.
4- اختبار مان – وتني U.
5- اختبار ولكوكسن T.
6- اختبار كروسكال – وليز.
7- اختبار فريدمان.

رابعاً: مقاييس العلاقة بين المتغيرات:

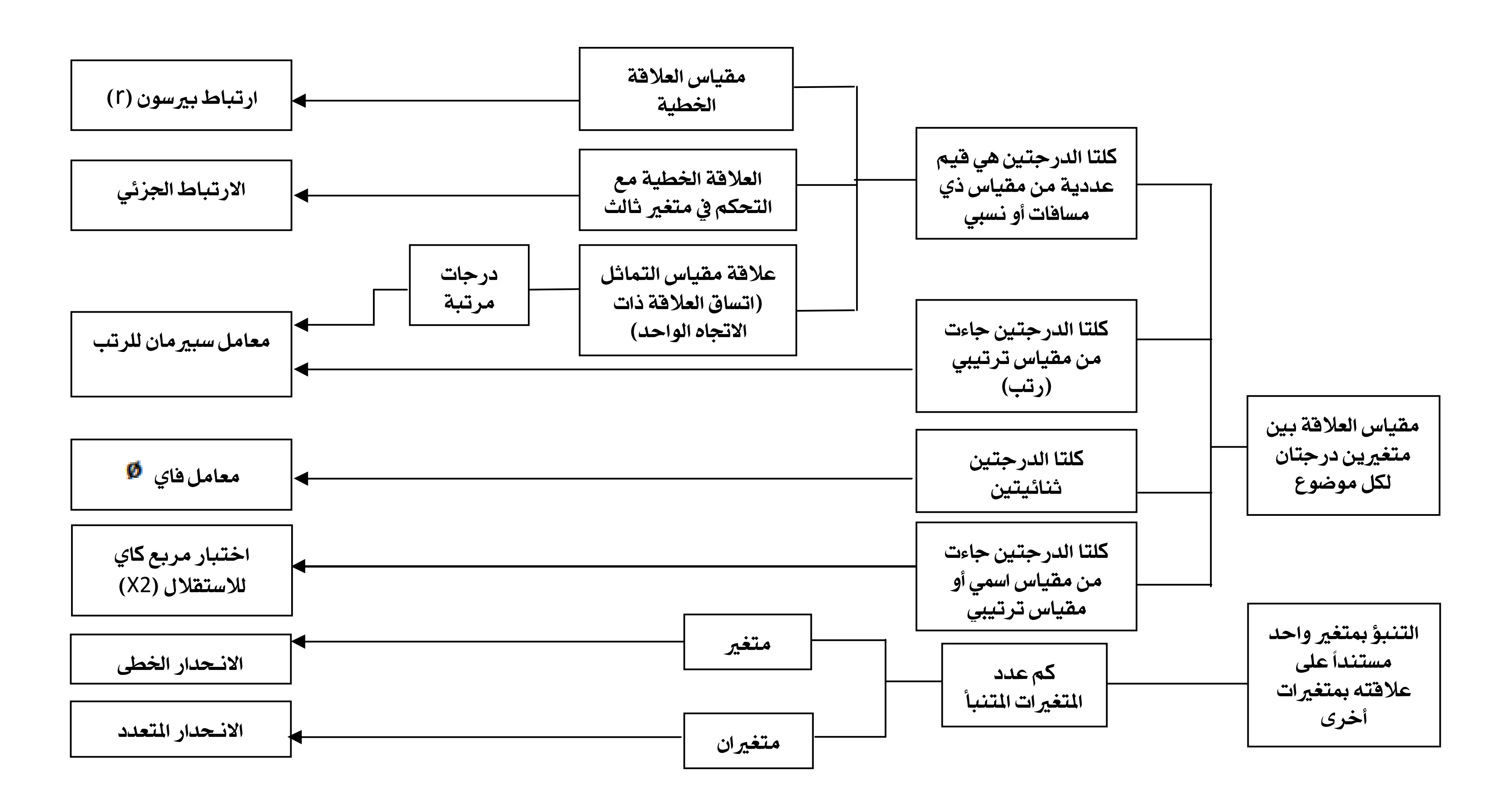
كل الطرق الإحصائية المختلفة التي تم وصفها في هـذا الجـزء يـراد منهـا أن تـستخدم نمطاً محدداً من البيانات. ولكي نـحدد الطريقة الملائمة، ينبغي على الباحث بادئ ذي بـدء أن يفحص بياناته ويعيّن ما هو نمط المتغير المستخدم، وما هـو مـستوى القيـاس المـستخدم لترميز المشاهدات.

وتشمل مقاييس العلاقة بين المتغيرات المقاييس التالية:

1- معامل بيرسون (r).
2- المعامل الجزئية $r_{yz.A}$.
3- معامل سبيرمان r_s.
4- معامل ϕ .
5- اختبار مربع كاي للاستقلال.
6- الانـحدار الخطي regression Line.
7- الانـحدار المتعدد Multiple Regression.

والشكل (4) خارطة قرار تساعد الباحث في اختيار الأسلوب الملائم لتقييم العلاقـة بين المتغيرات.

شكل (3-4): خارطة قرار تساعد الباحث في اختيار الأسلوب الملائم لتقييم العلاقة بين المتغيرات



ملحق (4)
بعض المعادلات الأساسية الواردة في الكتاب

1- متوسط مجتمع: بيانات مدرجة في قائمة: $u=\frac{\Sigma X}{N}$

N تشير إلى حجم المجتمع.

X تشير لكل درجة في التوزيع.

2- متوسط عينة: بيانات مدرجة في قائمة: = $\overline{X}=\frac{\Sigma X}{n}$

n تشير إلى حجم العينة.

3- متوسط عينة: بيانات تكرارية: $\overline{X}=\frac{\Sigma FX}{n}$

F تشير إلى التكرار لكل قيمة في التوزيع.

4- متوسط عينة: بيانات في فئات: $\overline{X}=\frac{\Sigma FM}{n}$

M تشير إلى مركز الفئة.

5- الانـحراف المعياري لمجتمع: بيانات مدرجة في قائمة: $\sigma=\sqrt{\frac{\Sigma(X-u)^2}{N}}$

6- الانـحراف المعياري لعينة: بيانات مدرجة في قائمة: $S=\sqrt{\frac{\Sigma(X-\overline{X})^2}{N-1}}$

$$S=\sqrt{\frac{\Sigma X^2-\frac{(\Sigma X)^2}{n}}{n-1}}$$

7- الانـحراف المعياري لعينة: بيانات تكرارية: $S=\sqrt{\frac{\Sigma FX^2-\frac{(\Sigma FX)^2}{n}}{n-1}}$

8- معامل التباين النسبي (CRV): $CRV = \frac{S}{\overline{X}} \times 100$

9- مؤشر التباين النوعي (IQV): $IQV = \frac{\text{فروق مشاهدة}}{\text{أقصى فروق محتملة}}$

K تشير إلى عدد الفئات. $\frac{n^2(K-1)}{2K}$ = أقصى فروق محتملة

10- درجة Z لوصف مجتمع: $Z = \frac{X-\mu}{\sigma}$

11- درجة Z لوصف عينة: $Z = \frac{X-\overline{X}}{S}$

12- لامبيدا Lambeda $\lambda = \frac{E_1 - E_2}{E_1}$

E_1 تشير إلى عدد الأخطاء بدون معلومات حول المتغير المستقل.

E_2 تشير إلى عدد الأخطاء بمعلومات عن المتغير المستقل.

13- جرامرز Cramer`s V: $V = \sqrt{\frac{X^2}{n(K-1)}}$

X^2 إحصاء مربع كاي لجدول التقاطع.

K عدد الصفوف أو عدد الأعمدة أيهما أصغر.

14- جاما GAMMA $G = \frac{N_C - N_d}{N_C + N_d}$

N_C تشير إلى عدد الأزواج المتوافقة.

N_d تشير إلى عدد الأزواج غير المتوافقة.

15- سومرز d: Somer`s d $d = \frac{N_C - N_d}{N_C + N_d + t_Y}$

t_y تشير إلى عدد الحالات المرتبطة على المتغير التابع مع التباين على المتغير المستقل.

16- كندل تاو – b: $tau-b=\frac{N_C-N_d}{\sqrt{(N_C+N_d+T_Y)(N_C+N_d+T_X)}}$

T_X تشير إلى عدد الحالات المرتبطة على المتغير المستقل مع التباين على المتغير التابع

17- كندل تاو - C: $tau-C=\frac{2K(N_C-N_d)}{N^2(K-1)}$

18- معامل سبيرمان للرتب: $r_S=1-\frac{6\Sigma D^2}{n(n^2-1)}$

19- معادلة الخط المستقيم. $Y=a\pm bX$

Y تشير إلى المتغير التابع.

X تشير إلى المتغير المستقل.

a كمية ثابتة وهي المساقة على محور (y) من البداية إلى النقطة التي تقطع محور (y) فهي قيمة (y) لقيمة (X) التي تساوي صفراً (0).

b تشير إلى ميل الخط.

+ تشير إلى التطابق الموجب.

– تشير إلى التطابق السالب.

20- معامل الانــحدار:

$$b=\frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}$$

$$b=\frac{N\Sigma(xy)-(\Sigma x)(\Sigma y)}{N\Sigma x^2-(\Sigma x)^2}$$

21- معامل ارتباط بيرسون (r):

$$r=\frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma\left[(x-\bar{x})^2\right]\left[(y-\bar{y})^2\right]}}$$

$$r=\frac{N\Sigma(xy)-(\Sigma x)(\Sigma y)}{\sqrt{\left[N\Sigma x^2-(\Sigma x)^2\right]\left[N\Sigma y^2-(\Sigma y)^2\right]}}$$

22- فترات الثقة للمتوسط:

$$\text{Lower Limit الحد الأدنى} = \overline{X} - Z\left(\frac{S}{\sqrt{n}}\right)$$

$$\text{upper Limit الحد الأعلى} = \overline{X} + Z\left(\frac{S}{\sqrt{n}}\right)$$

23- اختبار Z لمتوسط مفرد: $Z = \frac{\overline{x} - u}{\sigma / \sqrt{N}}$

24- اختبار t لمتوسط عينة واحدة: $= \quad t = \frac{\overline{X} - u}{S / \sqrt{N}}$

25- اختبار Z لنسبة ذي حدين:

حيث إن $Z = \frac{(P_S - 0.5) - P_u}{\sqrt{\frac{P_u(100 - P_u)}{n}}}$ $\quad P_S > P_u$

أو

حيث إن: $Z = \frac{(P_S + 0.5) - P_u}{\sqrt{\frac{P_u(100 - P_u)}{N}}}$ $\quad Pu > Ps$

$$\sigma P = \sqrt{P_u\left(\frac{1 - P_u}{n}\right)}$$

Pu تشير إلى نسبة المجتمع.

26- اختبار مربع كاي لحسن المطابقة والاستقلال $x^2 = \sum \frac{(F_O - F_e)^2}{F_e}$

Fo تشير إلى التكرار المشاهد.

Fe تشير إلى التكرار المتوقع.

df = (r-1) (C-1)

حيث إن: r تشير إلى عدد الصفوف

C تشير إلى عدد الأعمدة

27- اختبار t لمتوسطين متساويين:
$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sigma \overline{X} - \overline{X}}$$

(تقدير التباين المجمع)
$$\sigma\overline{\mathbf{x}} - \overline{\mathbf{x}} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

28- اختبار F لأكثر من متوسطين
$$F = \frac{\frac{SSB}{K-1}}{\frac{SSW}{n-K}}$$

$$TSS = SSB + SSW$$

$$TSS = \sum X - m\overline{X}^2$$

$$SSW = \sum (X - \overline{X}_S)^2$$

$$SSB = \sum N_S(\overline{X}_S - \overline{X})^2$$

$\overline{X}_s$ تشير إلى متوسط لعينة معطاة.

n_s تشير إلى عدد الحالات لعينة معطاة.

29- اختبار t لفرق المتوسط لعينتين مرتبطتين
$$t = \frac{\overline{X}_D}{S_D / \sqrt{n}}$$

$$\overline{X}_D = \frac{\Sigma D}{n}$$

$$S_D = \sqrt{\frac{\Sigma D^2 - \frac{(\Sigma D)^2}{n}}{N-1}}$$

30- اختبار مان – وتني U
$$UA + UB = n$$

$$U_A = n_V(n_B) + \frac{n_A(n_A + 1)}{2} - \sum R_A$$

$$U_B = n_A(n_B) + \frac{n_B(n_B +_1)}{2} - \sum R_B$$

التقريب الطبيعي لاختبار مان) وتني U
$$Z = \frac{x-u}{\sigma} = \frac{\frac{U - n_A n_B}{2}}{\sqrt{\frac{n_A n_B (n_A + n_B + 1)}{12}}}$$

31- اختبار كروسكال – وليز
$$H = \frac{12}{N(N+1)}\left(\sum \frac{T^2}{n}\right) - 3(N+1)$$

32- اختبار فريدمان
$$X_r^2 = \frac{12}{nk(k+1)}\sum R^2 - 3n(k+1)$$

33- مقاييس حجم التأثير
$$\text{Cohen`s d} = \frac{\text{فرق المتوسط}}{\text{الانحراف المعياري}}$$

(نسبة التباين التي تدخل في الاعتبار) and r^2 n^2

(لاختبار t المستقلة والمتكررة) $r^2 =$ $r_2 = \frac{t_2}{t_2 + df}$ أو Eta^2

(لاختبار مربع كاي للاستقلال) $phi\sqrt{\frac{x^2}{n}}$

$$\text{Cramer's V}\sqrt{\frac{x^2}{n(dF)}}$$

المصادر

أولاً: المراجع العربية.

1- جولي بالانت (ت) خالد العامري، التحليل الإحصائي باستخدام SPSS، دار الفاروق للنشر والتوزيع، مصر، 2009م.

2- عبد الجبار توفيق البياتي، البحث التجريبي واختبار الفرضيات في العلوم التربوية والنفسية، جهينة للنشر والتوزيع، عمان، 2006م.

3- عبد الله عامر الهمالي، أسلوب البحث الاجتماعي وتقنياته، ط3، منشورات جامعة قاريونس، 2003 م.

4- ـــــــــــــ، التقنيات الإحصائية ومناهج البحث، منشورات جامعة قاريونس، 2008 م.

5- ـــــــــــــ، التحديث الاجتماعي معالمه ونماذج من تطبيقاته، منشورات جامعة قاريونس، 2008م.

6- سعد اللافي، الإحصاء الاستنتاجي، ط 1، منشورات أكاديمية الدراسات العليا، طرابلس، 2003م.

7- محمد صبحي أبو صالح وعدنان محمد عوض، مقدمة في الإحصاء، مكتب الكتب الأردني، عمان، الأردن، 1990م.

8- مصطفى خلف عبد الجواد، الإحصاء الاجتماعي: المبادئ والتطبيقات، دار المسيرة للنشر والتوزيع والطباعة، الأردن، 2009م.

ثانياً : المراجع باللغة الإنجليزية :

1. Colin Robson, Real World Research , Black well Publishing , USA , 2005.
2. EARL BABBie , The basic of Social Research , Thomson Wadsworth , USA , 2005.
3. Frederick J. Gravetter and Larry B. Wallnau , Statistics for The Behavioral Sciences , 8th ed , Wadsworth Cengage Learning , USA , 2010.
4. George Argyrous , Statistics for Social and Health Research With a Guide to SPSS , Sage Publications , London , 2001.
5. George DiEkhoff, Statictics for the social and Behavioral Sciences: Univariat, Bivariate, Multivariate , WM. C. Brown Publications , USA , 1992.
6. Hubert Blalock, Jr. Social Statistics , Mcgraw Hill Company , Inc. , New york , 1972.
7. Hugh Coolicam and Hodder Arnold , Research Methods and Statistics in Psychology , 4 th ed , Hodder and Stoughton , Educational , London , 2004.
8. J. Richard Kendrick , Social Statistics: An Introduction using SPSS for Windows , 2th ed , Pearson Education , Inc. , 2005.
9. Joseph F. Healey , Statistics: A Tool for Social Research , Wadsworth Cengage Learning , 7th ed , USA , 2005.
10. ______________ , The Essentials of Statistics: A Tool for Social Research , Wadsworth Cengage Learning , USA , 2010.
11. N. J. Cohen , Statistical Power analysis for behavioral Science , SAN Diego , CA. Accademic Press , 1988.
12. R. Mark Sirikin , Statistics for Social Science , Sage Publications , International Oaks , London , 1995.
13. William E. Wagner , III , 2th ed , using SPSS for Social Statistics and Research Methods , Pine force Press , an Inprint of Sage , Publications , Inc. , USA , 2010.



المؤلف : أ. د. عبد الله عامر الهمالي

المؤهلات العلمية :

- ليسانس علم الاجتماع والفلسفة ، 1970 ، كلية الآداب ، الجامعة الليبية.
- ماجستير علم الاجتماع ، 1974 ، الولايات المتحدة .
- دكتوراه علم الاجتماع ، 1979 ، الولايات المتحدة .

الخبرة الإدارية :

- رئيساً لقسم علم الاجتماع - جامعة بنغازي (قاريونس سابقاً)، 1985-1994م
- عميداً لكلية الآداب - جامعة بنغازي (قاريونس سابقاً) ، 1994 - 1997 م
- مديراً للإدارة العامة للعلاقات العلمية والتعاون الفني بقطاع التعليم والبحث العلمي 1998 م .
- مديراً عاماً للإدارة العامة للبعثات الدراسية والتدريب بقطاع التعليم والتكوين المهني 1999-2000 م .
- أميناً لمركز البحوث والدراسات 2003-2008 م .
- نائباً لرئيس منظمة التقدم الدولية (سابقاً) فيينا-النمسا .

المؤلفات والبحوث العلمية :

- المرأة العربية والمشاركة السياسية : دراسة ميدانية (بالاشتراك) 1983م
- التغير الاجتماعي : أسسه وتطبيقاته (بالاشتراك) 1984 م
- التحديث الاجتماعي : معالمه ونماذج من تطبيقاته 1986 م
- أسلوب البحث الاجتماعي وتقنياته 2003 م
- التقنيات الإحصائية ومناهج البحث 2008 م
- إضافة إلى مجموعة بحوث علمية في مجال تخصصه منشورة في كتب ومجلات علمية متخصصة

الخبرة التدريسية :

- أستاذ علم الاجتماع بجامعة بنغازي والجامعات الليبية وخبرة تدريسية تزيد عن ثلاثين سنة في مجالات : مناهج البحث الاجتماعي ، تصميم البحوث الاجتماعية ، الإحصاء الاجتماعي ، التغير الاجتماعي والتحديث .

الندوات والمؤتمرات والملتقيات العلمية :

- شارك في العديد من الندوات والمؤتمرات والملتقيات العلمية بالجامعات ومراكز البحث العلمي بلغت أكثر من خمس وعشرين مشاركة .
- شارك في عدة لجان علمية وفنية بلغت في مجموعها أربعون لجنة .
- أشرف وناقش أكثر من خمس وثلاثين رسالة علمية (ماجستير ودكتوراه) في الجامعات الليبية والعربية .

هذه الصفحة توجد في خلفية الكتاب

هذا الكتاب

يهدف هذا الكتاب المرجعي في المقام الأول إلى تعلم الأساليب الإحصائية واستخدامها في تحليل البيانات . كما يتيح أيضا فرصة تساعد الدارسين والمهتمين بمجال البحث العلمي بشكل عام، والبحث في مجال العلم الاجتماعي بشكل خاص في فهم الأساليب الإحصائية الوصفية والاستدلالية واستخداماتها الملائمة في تحليل البيانات ، ومنحهم الثقة اللازمة للتعامل مع هذه الأساليب الإحصائية واستخداماتها بشكل علمي دقيق . إضافة إلى هذا كله ، يتضمن هذا الكتاب بعض الإجراءات المتعلقة باستخدام المجموعة الإحصائية في العلوم الاجتماعية (SPSS) ، والكيفية التي تتم بها عملية تفسير مخرجات هذه المجموعة الإحصائية .

يتألف هذا الكتاب من عشرين فصلا تغطي بدءا من الإحصاءات الوصفية وحتى الإحصاءات الاستدلالية المتقدمة . وينقسم الكتاب في ذلك إلى ستة أجزاء . وينتهي بمسرد لأهم المفاهيم الإحصائية الوصفية والاستدلالية وتعريفها ، ومنظم إحصائي يلخص الأساليب الإحصائية التي تم تغطيتها في هذا الكتاب .